

Santa Fe, New Mexico June 2nd - 4th, 2010









Contents

Agenda Overview 5
June 2 nd Agenda 7
Speaker Presentations (June 2 nd) 9
Genome Center Talks 25
Meet and Greet Party w/ Food & Beverages 34
Poster Session 35
June 3 rd Agenda 95
Speaker Presentations (June 3 rd) 97
Genome Finishing Talks 111
Tech Time Talks 117
Happy Hour(s) at Cowgirl BBQ w/ map 125
June 4 th Agenda127
Speaker Presentations (June 4 th) 129
Close of Meeting Discussion 141
Attendees 143
Maps & History of Santa Fe, NM 151

The 2010 "Sequencing, Finishing, Analysis in the Future" Organizing Committee:

- * Chris Detter, Ph.D., JGI-LANL Center Director, LANL
- * Johar Ali, Ph.D., Technology Development Team Leader, OICR
- * Ruby Archuleta, Genome Administrative Assistant, LANL
- * Patrick Chain, Metagenomics Team Leader, LANL
- * Michael Fitzgerald, Finishing Manager, Broad Institute
- * Bob Fulton, M.S., Sequence Improvement Group Leader, WashU
- * Darren Grafham, Finishing Coordinator, Sanger Institute
- * Jessica Hostetler, Genome Finishing and Analysis Manager, JCVI
- * Alla Lapidus, Ph.D., Finishing Coordinator Group Leader, LBNL-JGI
- * Donna Muzny, M.S., Director of Operations, BCM



PRODUCTS FOR NEXTGEN SEQUENCING



IDT can provide controls for your nextgen sequencing needs.

Single-stranded DNA Ultramers

- Up to 200 bases in length
- ESI mass spec QC for verification of full-length product
- PAGE purification available
- Ideal for highly repetitive constructs

Double-stranded DNA genes and miniGenes

- Sequence verification
- 2 µg of plasmid provided
- PCR priming and restriction sites easily included

Single-stranded RNA

- Up to 50 bases (longer targets individually quoted)
- HPLC purification available
- Provided in RNase-free buffer

miRNA cloning linkers for small RNA sequencing

- Based on pre-activated, adenylated linking technology
- Permits cloning from any RNA source in any species
- miRCat Small RNA Cloning Kit also available



06/02/201	10 - Wednesday	Abstract #	Title	Cucakar
7:30 - 8:30am	Breakfast	Abstract #	Title La Fonda Breakfast Buffet Scrambled Eggs with applewood smoked bacon and toast, Fruit jam and butter, Freshly brewed regular and decetfainered coffee.	х
8:30 - 8:45	Intro	x	decaffeinated coffee Welcome Intro from Los Alamos National Laboratory	Mary Neu
x	Session Chair	x	Session Chairs	Chair - Mike Fitzgerald
3:45 - 9:30	Keynote	FF0030	Complex Microbial Communities: We're not in Kansas anymore	Chair - Jessica Hostetle Dr. Claire Frasier- Liggett
9:30 - 9:50	Speaker 1	FF0004	Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis	Stephen Kingsmore
1:50 - 10:10	Speaker 2	FF0012	National Interagency Genome Science Coordinating Committee	Ron Walters
		FF0141	Retrospective View of the Response to Intentional Release of Pathogens	
10:10 - 10:30	Speaker 3	FF0141	Data Management Requirements for the Rapid Identification and Characterization of Unknown Genomic Samples	Nicole Rosenzweig
10:30 - 10:50 10:50 - 11:10	Break	FF0166	Beverages and snacks provided The Entire of Misrobial Conomics	X Nikos Kyrpides
	Speaker 4		The Future of Microbial Genomics	
1:10 - 11:30	Speaker 5 Speaker 6	FF0028	454 - Investing in the future generations - from sequencing symbiotic environments to Peer-Reviewed Online Science Education	Clotilde Teiling Haley Fiske
11:30 - 11:50 1:50 - 12:10	Speaker 7	FF0069 FF0025	illumina - New Applications and Sequencing Instrumentation SOLID - The \$6,000 Genome and beyond	Michael Rhodes
	.,		Coronado Lunch Buffet Tossed mixed greens with vegetable garnishes served with a choice of dressing,	
12:10 – 130pm	Lunch	RainDance	Grilled chicken breast with spicy barbeque vinaigrette, Three cheese tortellini with classic marinara and Alfredo sauces Fresh seasonal vegetables, Chef's choice of starch, Sourdough rolls and butter, Beverages	Sponsored by RainDanc
t	Session Chair	x	Session Chairs	Chair - Bob Fulton Chair - Patrick Chain
1:30 – 1:50	Speaker 8	FF0065	Pacific Biosciences - Towards Autofinishing in Complex Genomes: Hybrid Assembly of Rhodopseudomonas palustris Without	Steve Turner
	1,		Manual Intervention Qdot - Single molecule real-time nucleic acid sequencing-by-synthesis using Quantum-dot (Qdot) nanocrystal DNA polymerases with	
:50 – 2:10	Speaker 9	FF0124	FRET-based detection	Peter Vander Horn
2:10 - 3:10	Panel Discussion	x	Next Generation Sequencing Technology Panel Discussion	Chair - Chad Nusbaum
3:10 – 3:40	Break	FF0060	Beverages and snacks provided Genome Sequencing Center at NCGR	x Faye Schilkey
		FF0074	Genome Sequencing Center at NCGK Genome Sequencing Center at DOE JGI	Chris Daum
		FF0051	Genome Sequencing Center at Sanger	Darren Grafham
0.40 5.00	Tech Time Talks	FF0125	Genome Sequencing Center at Broad	Mike Fitzgerald
:40 - 5:20pm	(10 min each) with question panel	FF0129	Genome Sequencing Center at Baylor	Donna Muzny
		FF00101	Genome Sequencing Center at WashU	Bob Fulton
		FF0085 FF0227	Genome Sequencing Center at JCVI	Jessica Hostetler
	Postera "	. 1 0221	Genome Sequencing Center at OICR	Johar Ali
i:30 – 7:00pm	Posters - even #s Meet & Greet Party	EVEN #s	Poster Session with Meet & Greet Paty (Sponsorted by Roche) Food & Drinks	Sponsored by Roche 5:30pm- 9:00pm
7:00 - 8:30pm	Posters - Odd #s	ODD #s	Poster Session with Meet & Greet Party (Sponsorted by Roche) Food & Drinks	Sponsored by Roche
	Meet & Greet Party		-, -, -, -, -, -, -, -, -, -, -, -, -, -	5:30pm- 9:00pm
06/03/20 Time	010 - Thursday Type	Abstract #	Title	Speaker
7:30 - 8:30am	Breakfast	x	Santa Fe Breakfast Buffet Frittata with applewood smoked bacon and English muffin, Fruit jam and butter, Freshly brewed regular	x
1.20 0.4E	Intro	v	and decaffeinated coffee	Jim Bristow
1:30 - 8:45	Intro	X	Welcome Back from DOE Joint Genome Institute	Chair - Alla Lapidus
	Session Chair	x	Session Chairs	Chair - Mike Fitzgeralc
:45 - 9:30	Keynote	FF0003	DNA Sequencing Technology as the Engine of Scientific Advancement	Dr. Chad Nusbaum
:30 - 9:50	Speaker 1	FF0044	Large Genome Assembly with 454 Data	Jim Knight
:50 -10:10	Speaker 2	FF0201	A General Method for Assembling Genomes from Illumina data	Dariusz Przybylski
0:10 - 10:40	Break	x	Beverages and snacks provided	x
10:40 -11:00	Speaker 3	FF0172	Rnnotator: ade novo transcriptome assembly pipeline from short RNA-Seq reads	Jeff Martin
11:00 - 11:20	Speaker 4	FF0006	Phytophthora capsici: genome assembly of a polymorphic organism	Joann Mudge
1:20 - 11:40	Speaker 5	FF0165	Breaking the BLAST bottleneck	Robert Edgar
11:40 - 12:00	Speaker 6	FF0061	Towards experimental gene annotation by deep sequencing	Andrew Bradbury
12:00 - 1:30pm	Lunch	illumina	New Mexican Lunch Buffet Green salad with vegetable garnishes and assorted dressings, Beef fajita quesadilla garnished with lettuce, tomato, sour cream and guacamole, Cheese enchiladas with red chile, Black beans and rice,	Sponsored by illumina
			Beverages	Chair - Donna Muzny
· ·	Session Chair	x	Session Chairs	Chair - Johar Ali
1:30 - 1:50	Speaker 7	FF0116	The Medicago truncatula HapMap Project: Resequencing of 400Medicago Genotypes and 48Sinorhizobium strains using Illumina's 2 ^d Generation solexa Technology	Arvind Bharti
1:50 - 2:10	Speaker 8	FF0111	Sequencing the unrearranged human immunoglobulin heavy chain locus from	Rene Warren
1:50 - 2:10	Speaker 8	FFUIII	hydatidiform mole	kene warren
2:10 - 2:30	Speaker 9	FF0098	Implementation of a cost-effective pyrosequencing process through automation, molecular barcoding and process improvement	Niall Lennon
		FF0040	Univ. of Florida - Use of optical mapping in bacterial genome finishing	Dibyendu Kumar
	Genome Finishing (Selected Posters)	FF0105	Baylor Finishing - Microbe Assembly Strategies and Finishing at BCM-HGSC	Christian Buhay
2:30 - 3:30pm	(10 min each)	FF0100	Sanger Finishing - Gold Standard: Finishing Using Next Generation Technologies	Andries Van Tonder
	with question panel	FF0127 FF0119	Broad Finishing - Multiplex Template Preparation for Finishing With Next Generation Sequencing LANL Finishing - A high-throughput next generation genome finishing pipeline	Anna Montmayeur Karen Davenport
1:30 - 3:50	Break	х	Beverages and snacks provided	x
		FF0209	Caliper - Advanced Nucleic Acid Fractionation for Next Generation Sequencing Sample Preparation	Mark Roskey
		FF0090	Covaris - AFA-based DNA Fragmentation: A Critical First Step for the Optimization of Next Gen Sequencing	Jim Laugharn
	Took Time Taller	FF0164	OpGen - Use of Optical Mapping to Aid in the Assembly and Finishing of Human Microbiome Genome Projects	Trevor Wagner
:50 - 5:20pm	Tech Time Talks (15 min each)	FF0035	Geospiza - Increasing the Scale of Deep Sequencing Data Analysis with BioHDF	Todd Smith
		FF0193	CLC bio - Exploring highly specialized Next Generation Sequencing Analysis tools in CLCbio's Genomics Workbench	Saul Kravitz
		FF0059	NABsys - Hybridization Assisted Nanopore Sequencing	John Oliver
i:45 - 7:45	Happy Hour	LifeTech	Happy Hour at Cowgirls Cafe - Sponsored by Life Technologies (3 drink tickets per person)	Map Will be Provided
:45 - bedtime	on your own	x	Dinner and night on your own - enjoy	x
06/04/2	2010 - Friday			
Time	Туре	Abstract #	Title	Speaker
7:30 - 8:30am	Breakfast	x	Breakfast Buffet Harvey House French toast with warm maple syrup and butter, Sausage links, Freshly brewed regular and decaffeinated coffee	x
1:30 - 8:45	Intro	FF0232	Welcome Back - Review Genome Standards Paper What's next?	Patrick Chain
	Session Chair	x	Session Chairs	Chair - Patrick Chain
1:45 - 9:30	Keynote	FF0048	Analysis of metagenomic sequences, from megabases to terabases	Chair - Darren Grafharr Dr. Folker Meyer
:30 - 9:50	Speaker 1	FF0142	Community genomic analysis of relatively low diversity ecosystems: opportunities and challenges for up-scaling	Jill Banfield
	Speaker 2	FF0115	Whole genome shotgun metagenomic sequencing, assembly, and analysis of human microbial communities associated with health and disea	seJoe Petrosino
9:50 - 10:10				
		x	Beverages and snacks provided Grepping Life: A New Paradigm for Analyzing Metagenomic Data	Joel Berendzen
0:10 - 10:30	Break	EE0249	отерриту спо. л tvew maradigiti for Analyzing inetagenomic Data	
0:10 - 10:30 0:30 - 10:50	Speaker 3	FF0218	One Microbial Cell. One Complete Genome	Tania Woyke
0:10 - 10:30 0:30 - 10:50 0:50 - 11:10	Speaker 3 Speaker 4	FF0160	One Microbial Cell, One Complete Genome Single Amplified Whole Genome Shotour - Applications	Tanja Woyke
0:10 - 10:30 0:30 - 10:50 0:50 - 11:10 1:10 - 11:30	Speaker 3 Speaker 4 Speaker 5	FF0160 FF0135	Single Amplified Whole Genome Shotgun – Applications	Eric Pelletier
0:10 - 10:30 0:30 - 10:50 0:50 - 11:10 1:10 - 11:30 1:30 - 11:50	Speaker 3 Speaker 4 Speaker 5 Speaker 6	FF0160 FF0135 FF0208	Single Amplified Whole Genome Shotgun – Applications Whole Genome Sequencing of Individual Bacterial Cells from the Ocean	Eric Pelletier Rex Malmstrom
0:10 - 10:30 0:30 - 10:50 0:50 - 11:10 1:10 - 11:30 1:30 - 11:50	Speaker 3 Speaker 4 Speaker 5	FF0160 FF0135	Single Amplified Whole Genome Shotgun – Applications	Eric Pelletier
	Speaker 3 Speaker 4 Speaker 5 Speaker 6 Speaker 7 Closing Discussions	FF0160 FF0135 FF0208	Single Amplified Whole Genome Shotgun – Applications Whole Genome Sequencing of Individual Bacterial Cells from the Ocean DNA sequencing from single bacterial cells using the multiple displacement amplification (MDA) reaction Closing Discussions - discuss next year's meeting	Eric Pelletier Rex Malmstrom
0:10 - 10:30 0:30 - 10:50 0:50 - 11:10 1:10 - 11:30 1:30 - 11:50 1:50 - 12:10	Speaker 3 Speaker 4 Speaker 5 Speaker 6 Speaker 7	FF0160 FF0135 FF0208	Single Amplified Whole Genome Shotgun – Applications Whole Genome Sequencing of Individual Bacterial Cells from the Ocean DNA sequencing from single bacterial cells using the multiple displacement amplification (MDA) reaction	Eric Pelletier Rex Malmstrom Roger Lasken

06/02/201	0 - Wednesday			
Time	Туре	Abstract #	Title	Speaker
7:30 - 8:30am	Breakfast	x	La Fonda Breakfast Buffet Scrambled Eggs with applewood smoked bacon and toast, Fruit jam and butter, Freshly brewed regular and decaffeinated coffee	x
8:30 - 8:45	Intro	x	Welcome Intro from Los Alamos National Laboratory	Mary Neu
x	Session Chair	х	Session Chairs	Chair - Mike Fitzgerald Chair - Jessica Hostetler
8:45 - 9:30	Keynote	FF0030	Complex Microbial Communities: We're not in Kansas anymore	Dr. Claire Frasier- Liggett
9:30 – 9:50	Speaker 1	FF0004	Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis	Stephen Kingsmore
9:50 - 10:10	Speaker 2	FF0012	National Interagency Genome Science Coordinating Committee Retrospective View of the Response to Intentional Release of Pathogens	Ron Walters
10:10 – 10:30	Speaker 3	FF0141	Data Management Requirements for the Rapid Identification and Characterization of Unknown Genomic Samples	Nicole Rosenzweig
10:30 – 10:50	Break	х	Beverages and snacks provided	x
10:50 – 11:10	Speaker 4	FF0166	The Future of Microbial Genomics	Nikos Kyrpides
11:10 – 11:30	Speaker 5	FF0028	454 - Investing in the future generations - from sequencing symbiotic environments to Peer-Reviewed Online Science Education	Clotilde Teiling
11:30 – 11:50	Speaker 6	FF0069	illumina - New Applications and Sequencing Instrumentation	Haley Fiske
11:50 - 12:10	Speaker 7	FF0025	SOLiD - The \$6,000 Genome and beyond	Michael Rhodes
12:10 – 130pm	Lunch	RainDance	Coronado Lunch Buffet Tossed mixed greens with vegetable garnishes served with a choice of dressing, Grilled chicken breast with spicy barbeque vinaigrette, Three cheese tortellini with classic marinara and Alfredo sauces Fresh seasonal vegetables, Chef's choice of starch, Sourdough rolls and butter, Beverages	Sponsored by RainDance
x	Session Chair	x	Session Chairs	Chair - Bob Fulton Chair - Patrick Chain
1:30 – 1:50	Speaker 8	FF0065	Pacific Biosciences - Towards Autofinishing in Complex Genomes: Hybrid Assembly of Rhodopseudomonas palustris Without Manual Intervention	Steve Turner
1:50 – 2:10	Speaker 9	FF0124	Qdot - Single molecule real-time nucleic acid sequencing-by-synthesis using Quantum-dot (Qdot) nanocrystal DNA polymerases with FRET-based detection	Peter Vander Horn
2:10 – 3:10	Panel Discussion	x	Next Generation Sequencing Technology Panel Discussion	Chair - Chad Nusbaum
3:10 – 3:40	Break	x	Beverages and snacks provided	x
	(10 min each) with question panel	FF0060	Genome Sequencing Center at NCGR	Faye Schilkey
3:40 - 5:20pm		FF0074	Genome Sequencing Center at DOE JGI	Chris Daum
		FF0051	Genome Sequencing Center at Sanger	Darren Grafham
		FF0125	Genome Sequencing Center at Broad	Mike Fitzgerald
		FF0129	Genome Sequencing Center at Baylor	Donna Muzny
		FF00101	Genome Sequencing Center at WashU	Bob Fulton
		FF0085 FF0227	Genome Sequencing Center at JCVI	Jessica Hostetler
			Genome Sequencing Center at OICR	Johar Ali
5:30 – 7:00pm	Posters - even #s Meet & Greet Party	EVEN #s	Poster Session with Meet & Greet Party (Sponsorted by Roche) <u>Food & Drinks</u>	Sponsored by Roche 5:30pm- 9:00pm
7:00 - 8:30pm	Posters - Odd #s Meet & Greet Party	ODD #s	Poster Session with Meet & Greet Party (Sponsorted by Roche) Food & Drinks	Sponsored by Roche 5:30pm- 9:00pm

Speaker Presentations (June 2nd) Abstracts are in order of presentation according to Agenda

FF0030

Keynote

Claire Frasier- Liggett

Institute for Genome Sciences, University of Maryland School of Medicine

Complex Microbial Communities: We're not in Kansas anymore

NOTES

Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis

Sergio E. Baranzini1, Joann Mudge2, Jennifer C. van Velkinburgh2, Pouya Khankhanian1, Irina Khrebtukova3, Neil A. Miller2, Lu Zhang3, Andrew D. Farmer2, Callum J. Bell2, RyanW.Kim2, Gregory D.May2, Jimmy E.Woodward2, Stacy J. Caillier1, Joseph P. McElroy1, Refujia Gomez1, Marcelo J. Pando4, Leonda E. Clendenen2, Elena E. Ganusova2, Faye D. Schilkey2, Thiruvarangan Ramaraj2, Omar A. Khan5, Jim J. Huntley3, Shujun Luo3, Pui-yan Kwok6,7, Thomas D. Wu8, Gary P. Schroth3, Jorge R. Oksenberg1,7, Stephen L. Hauser1,7 & Stephen F. Kingsmore2

1Department of Neurology, University of California at San Francisco, San Francisco, California 94143, USA. 2National Center for Genome Resources, Santa Fe, New Mexico 87505,USA. 3Illumina Inc., Hayward, California 94545, USA. 4Stanford Medical School Blood Center, Palo Alto, California 94303, USA. 5Department of Neurology, Wayne State Medical School, Detroit, Michigan 48201, USA. 6Cardiovascular Research Institute, University of California at San Francisco, California 94143, USA. 7Institute for Human Genetics, University of California at San Francisco, San Francisco, California 94143, USA. 8Department of Bioinformatics, Genentech Inc., South San Francisco, California 94080, USA.

Monozygotic or 'identical' twins have been widely studied to dissect the relative contributions of genetics and environment in human diseases. In multiple sclerosis (MS), an autoimmune demyelinating disease and common cause of neurodegeneration and disability in young adults, disease discordance in monozygotic twins has been interpreted to indicate environmental importance in its pathogenesis1-8. However, genetic and epigenetic differences between monozygotic twins have been described, challenging the accepted experimental model in disambiguating the effects of nature and nurture9-12. Here we report the genome sequences of one MS-discordant monozygotic twin pair, and messenger RNA transcriptome and epigenome sequences of CD41 lymphocytes from three MS-discordant, monozygotic twin pairs. No reproducible differences were detected between co-twins among 3.6 million single nucleotide polymorphisms (SNPs) or 0.2 million insertion-deletion polymorphisms. Nor were any reproducible differences observed between siblings of the three twin pairs in HLA haplotypes, confirmed MS-susceptibility SNPs, copy number variations, mRNA and genomic SNP and insertion-deletion genotypes, or the expression of 19,000 genes in CD41 T cells. Only 2 to 176 differences in the methylation of 2 million CpG dinucleotides were detected between siblings of the three twin pairs, in contrast to 800 methylation differences between T cells of unrelated individuals and several thousand differences between tissues or between normal and cancerous tissues. In the first systematic effort to estimate sequence variation among monozygotic cotwins, we did not find evidence for genetic, epigenetic or transcriptome differences that explained disease discordance. These are the first, to our knowledge, female, twin and autoimmune disease individual genome sequences reported.

National Interagency Genome Science Coordinating Committee Retrospective View of the Response to Intentional Release of Pathogens

Ronald A. Walters, Ph.D.

Pacific Northwest National Laboratory Richland, Washington, USA

The September-October 2001 releases of anthrax spores shortly after the 9/11 attack on the World Trade Center and Pentagon were sobering in their timing, the potential for great harm, and their impact on our society and global view. That remains true today. A very recent report from the Congressional Research Service summarized the issues associated with the federal response to threats of bioterrorism. The continued attempts to launch terror attacks on U.S. citizens are a clear indication that the threat has not disappeared. Not surprisingly, there has been both support of and criticism for every federal response taken to address the threat of an intentional release of pathogens, and it is not the intent here to debate outcomes that have led to the current infrastructure to deal with dangerous pathogens associated with biocrime or bioterrorism. What will be presented is a description of and success in mobilizing the federal agencies that could assist in characterizing the genomes of dangerous pathogens, information for which was sorely lacking in 2001. The National Interagency Genome Sciences Coordinating Committee (NIGSCC) was established in 2001 to coordinate the sequencing of pathogens considered to be the most dangerous agents of potential biocrime or bioterrorism. The NIGSCC was chaired by Dr. Rita Colwell, Member organizations originally contributing representatives then NSF Director. included NSF, NIH/NIAID, IC, DOE, DOJ, CDC, USDA, DOD, and DHS. The NIGSCC continues to function and has added additional representatives from FDA, EPA, NBACC, DTRA, and various other DOD organizations. Accomplishments and perspectives will be presented.

Data Management Requirements for the Rapid Identification and Characterization of Unknown Genomic Samples

 $\underline{\text{C. Nicole Rosenzweig}}^{1,2}, \text{ Alvin Liem}^{1,2}, \text{ Ed Fochler}^{1,2}, \text{ Greg Donarum}^{1,2}, \text{ Jessica Hill}^{1,2}, \text{ Lauren A. McNew}^{1,3}, \text{ Henry S. Gibbons}^{1}, \text{ Evan W. Skowronski}^{1}$

The rapid identification and characterization of uncharacterized or unknown samples following shotgun sequencing requires a rigorous data management plan. To support this end, Edgewood Chemical and Biological Center (ECBC) Genomics IPT has developed an informatics infrastructure to support the objective of identifying emerging and genetically engineered bacterial pathogens. An automated analytical pipeline has been developed and is required in order to (1) scale the sequencing effort to develop a capability to filter, analyze and interpret a large amount of data, and (2) to ensure standardized analytical methodologies are pursued while capturing and documenting the exact methodology used at any time. The analytical pipeline identifies the main genomic components of a sample, identifies nearest neighbors, and enables research-driven analysis.

Samples are sequenced using the Roche 454 GS FLX Genome Sequencers using the Titanium Reagent Platform. A Laboratory Information Management System (LIMS) captures all laboratory data. Once sequencing is complete, the automated pipeline is initiated. Roche De Novo Assembler and Reference Mapper, MPI BLAST, and MUMmer are utilized in analysis. Analytical results are parsed back into LIMS. When a species-level match is found in the reference database, the unmapped reads are reevaluated to determine if plasmids, phages, or other genomic components are present.

Currently, purified bacterial samples are processed at ECBC. Analytical pipelines are in place to support this analysis. All samples with species-level reference genomes available in the Genbank RefSeq database were automatically identified correctly. Additionally, results suggest the ability to identify the presence of insertion elements, plasmids, and the main chromosomal components of the samples. These results suggest that automated analysis can provide an initial identification and characterization of an unknown sample following whole genome sequencing.

Support for this project was provided by Defense Threat Reduction Agency project TMT_CB2847 to H.S.G. and E.W.S.

¹Genomics IPT, Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD ²OptiMetrics, Inc., ³Battelle Memorial Institute

The Future of Microbial Genomics

Nikos C. Kyrpides

Genome Biology Program, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

At the time of the completion of 1000 microbial genomes, the field is poised at a crossroads. The enabling technology which birthed the field and brought it to this point is fundamentally different from the one that is currently pushing it to the next phase. The future holds great promise for far-reaching advancements in microbiology as well as in diverse related sciences. To realize that potential will require meeting the challenges that have accompanied the rapid development of the underlying technology and the exponential growth of data. New technologies provide unprecedented opportunities, but also call for conceptual shifts. Experience gained in the first decade of genomics can guide the improved approaches now needed for the selection of genome sequencing projects and their funding, for genome annotation standards, as well as for data analysis and integration. I will discuss the current challenges in the field and present a vision of its future.

NOTES

NOTES

Investing in the future generations - from sequencing symbiotic environments to Peer-Reviewed Online Science Education"

Clotilde Teiling

454 Genome Sequencing, Roche Applied Science

While historically, symbiosis has received less attention than other interactions, it has become a booming field in the last decades as we are becoming more aware of our environment. Symbiosis between species is increasingly recognized to be an important selective force behind evolution. As high-through put sequencing is more available, sequencing one genome is not sufficient anymore, one genome with its entire environment gives a lot more information on how it is affected by disease, environmental desiquilibrium, etc...

Roche applied Science serves the scientific research community with an extensive product portfolio featuring cutting-edge instruments and reagents for genome sequencing, micro array analysis, nucleic acid purification, real-time PCR and cell analysis. But we also provide a wealth of valuable resources from educational websites to dedicated technical teachers, showing a commitment to today's future generation.

New Applications and Sequencing Instrumentation

Haley Fiske

illumina, Inc. 25861 Industrial Blvd, Hayward, CA 94545

Improvements to the Illumina platforms and methods will be discussed. Illumina has launched two new sequencing platforms. These platforms will be briefly discussed. Method improvements will cover whole transcriptome sequencing of samples from small amounts of nucleic acid and ffpe. The effect of PCR on library amplification will also be covered.

The \$6,000 Genome and Beyond

Michael D. Rhodes

Life Technologies, 500 Cummings Center, Beverly, MA 01915

The amazing growth in the capabilities of Sequencing Platforms has outstripped Moore's law over the past few years, both in terms of increases in throughput and decreases in cost of sequencing. Life Technologies is the world's leading sequencing company and has a complete portfolio of sequencing solutions from a single capillary machine through to ultra high throughput sequencing, with single molecule solutions available soon.

This growth in throughput and range of solutions has made the decision on what solution to select challenging. With the ability to easily generate enormous volumes of data the challenges have moved to workflow and data analysis. The intellectual challenge becomes the process of designing and constructing libraries so that the desired information is sequenced and then extracting that data from the resultant sequences.

In this talk, various technologies will be presented, including the SOLiD™ system and Single Molecule Sequencing. The SOLiD 4 system generates up to 100 gigabases of mapped data in a single run, some of the applications successfully carried out on the SOLiD system include:

Complete cancer genome re-sequencing, single cell transcriptome analysis, Methylation and de Novo sequencing. The latest applications will be presented including results and data analysis and automation solutions.

The SOLiD instrument is For Research Use Only, and is not intended for any animal or human therapeutic or diagnostic use

Lunch

12:10 - 1:30pm

Sponsored by



Towards Autofinishing in Complex Genomes: Hybrid Assembly of *Rhodopseudomonas palustris* Without Manual Intervention

Steve Turner

Pacific Biosciences

Despite their nearly universal use in DNA sequencing for several decades, DNA polymerases possess performance characteristics far beyond what first and second generation sequencing technologies have achieved. Through the use of phospholinked nucleotides, which can be incorporated with near-native kinetics by DNA polymerases, and zero-mode waveguides, which allow single-molecule detection at the high nucleotide concentrations required for DNA replication, the natural power of these enzymes has been harnessed for single-molecule, real-time (SMRT) DNA sequencing. Here, the high processivity of certain phage polymerases is translated into long readlength and applied to the shotgun whole genome assembly of a hydrogenproducing bacterium, Rhodopseudomonas palustris (R. pal). To demonstrate the effectiveness of long reads and strobe reads in resolving complex genome structure, we employed a hybrid assembly strategy using Pacific Biosciences' long reads and strobe reads in conjunction with reads from the Illumina sequencing platform. An important component of our approach was the development of fast and sensitive methods for overlapping SMRT reads with assembled contig sequences and In conjuction with this novel overlap algorithm we developed an themselves. automated pipeline for combining Illumina reads and SMRT reads through the steps of filtering, assembly, and scaffolding. The resulting assembly produced significantly fewer contigs and longer contig lengths than the starting Illumina assembly, with no misassembles when compared to the finished genome sequence. We show the role these assemblies play in the network analysis of hydrogen production by R. pal.

Single molecule real-time nucleic acid sequencing-by-synthesis using Quantum-dot (Qdot®) nanocrystal DNA polymerases with FRET-based detection

Peter Vander Horn, Joseph M. Beechem and Manfred W. Lee

Genetic Systems, Life Technologies, Carlsbad CA USA

Our Generation-1 (Sanger-based) and Generation-2 (SOLiD) sequencing platforms have helped to define the limits of current DNA sequencing technology. continuing efforts to advance (& complement) our Gen-1 and Gen-2 technologies, Life Technologies has developed a single molecule real-time sequencing-by-synthesis technology that builds upon our core competencies in Qdot® nanocrystals, Molecular Probes[®] organic dyes, Invitrogen™ DNA polymerases, and Applied Biosystems experience in manufacturing robust DNA sequencing instrumentation. component of our Gen-3 platform is the complete "portability" of the DNA sequencing engine. The sequencing engine is (essentially) a reagent, ~ 10nm in size, capable of sequencing DNA on a variety of substrates, from simple microscope coverslips, to nanochannels, to complex surfaces. Five-color fluorescence resonance energytransfer technology (FRET) (donor signal plus 4 acceptor signals), forms the basis of detection, as a single excitation line excites a donor-labeled DNA polymerase. The extra "fifth-color" (from the donor-polymerase) provides a very specific "anticorrelationsignal" (donor-down signals temporally linked to acceptor-up base-calling signals), that allows for more accurate base-calling. Ultimately, sequencer portability enables reagent-exchange, which in turn yields a highly flexibly sequencing capable of achieving tunable continuous read-lengths and accuracies. In one sequencing mode, when the single-molecule sequencers "wear-out" in mid-run (e.g., after sequencing 1Kb or so), new Qdot® polymerases can be added and the sequencing can be continued from exactly where the previous sequencer left-off. In this manner, the read-length can be continuously extended without requiring circularization of genomic templates. In another sequencing mode, both the sequencer and newly synthesized template can be washed out, new Qdot® polymerases can be added, and the template can then be resequenced (termed recursive sequencing), greatly increasing the net accuracy. In this manner, one can "tune" the desired accuracy one wishes to obtain by adjusting the number of reagent exchanges. A detailed description of the step-by-step construction and implementation of this portable single molecule sequencing engine will be Examples of real-time sequencing of homopolymeric, patterned, and presented. complex templates will be shown.

Panel Discussion Notes

Panel Discussion Notes

The Genome Sequencing Center at NCGR

Ryan W. Kim, Jimmy E. Woodward, Peter B. Ngam and Faye D. Schilkey

National Center for Genome Resources, Santa Fe, NM

National Center for Genome Resources (NCGR) is a non-profit research institute whose mission is to improve human health and nutrition through genome sequencing and analysis. The Genome Sequencing Center at NCGR offers next generation sequencing, genotyping, and analysis services, currently housing eight Illumina Genome Analyzers IIx for sequencing and Illumina's BeadExpress and iScan for genotyping. Sequencing services include whole genome and transcriptome shotgun sequencing, ChIP, small RNA, whole or targeted exome, and methylome sequencing. This summer, NCGR is one of only 10 sites including the Broad institute of MIT and Harvard, CSHL, JGI, and Wash U who are receiving the third generation Pacific Biosystem single molecule sequencing instrument. The PacBio system boasts very long nucleotide read lengths that provide fidelity especially helpful for de novo sequencing and resequencing projects.

NCGR has a longstanding reputation for developing effective bioinformatics tools for scientific research evident by two recent awards: 2009 Bio-IT World Best Practices in Basic Research in Schizophrenia and 2009 Laureate by Computerworld for Alpheus® a web-based analysis pipeline for variant and expression detection of next generation sequencing data.

Leveraging NCGR's own research, sequencing and analysis experience, the NCGR Genome Sequencing Center provides an unprecedented knowledgebase and facility with approximately 100 customers/collaborators to date from around the globe. Please contact Faye D. Schilkey at seq@ncgr.org to find out how NCGR can accelerate your research.

An inside look at the JGI's Sequencing Operation

Christopher Daum and Susan Lucas

US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA

The U.S. Department of Energy (DOE) Joint Genome Institute's (JGI) Production Sequencing group is committed to the generation of high-quality genomic DNA sequence to support the mission areas of renewable energy generation, global carbon management, and environmental characterization and clean-up. Within the JGI's Production Sequencing group, a robust Roche Genome Sequencer and Illumina Genome Analyzer pipeline has been established. Optimization of these sequencer pipelines has been ongoing with the aim of continual process improvement of the laboratory workflow. Discussion will include assessment of gDNA through data analysis, including assessment of quality on the production line.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

LLNL-ABS-408975 LBNL-1310E Abs.

Sequencing at the Wellcome trust Sanger institute

Darren Grafham

The Wellcome Trust Sanger Institute

The Wellcome Trust Sanger Institute (WTSI) has been a world leader in genomic sequencing for over 15 years and supports over 35 faculty researchers at the institute. The institute currently has 3 platforms for sequencing: 3730 capillary, 454 FLX and Illumina GAII which provides a diverse range of projects and needs. A brief overview of the variety of projects supported along with the lims, software and pipelines available will be presented.

Center Presentations: NexGen Production Sequencing and Analysis

Mike Fitzgerald

Broad Institute

Center Presentations: NexGen Production Sequencing and Analysis

Donna Muzny

Baylor College of Medicine

FF0101, #1

Center Presentations: NexGen Production Sequencing and Analysis

Bob Fulton

The Genome Center at Washington University, 4444 Forest Park Blvd., Saint Louis, MO 63108

FF0085, #1

Center Presentations: NexGen Production Sequencing and Analysis

Jessica Hostetler

The J Craig Venter Institute, Rockville, MD, U.S.A.

Center Presentations: NexGen Production Sequencing and Analysis

Johar Ali

The Ontario Institute for Cancer Research (OICR)

Panel Discussion Notes

Meet and Greet Party

600pm - 830pm, June 2nd

Sponsored by Roche Diagnostics

Enjoy!!!



Poster Presentations (June 2nd) Even #'s 5:30-7:00pm, Odd #'s 7:00-8:30pm

FF0024

12x coverage 454/Sanger Hybrid Assembly of Cucumber (*Cucumis sativus* L. cv. Borszczagovski) genome – the most efficient way to begin post genomic era

Woycicki R., Witkowicz J., Pawelkowicz M, Siedlecka E., Gutman W., Plader W., Seroczynska A., Smiech M., Niemirowicz-Szczytt K., Karpinski S, Malepszy S., Przybecki Z.

Dept. of Plant Genetics, Breeding & Biotechnology, Faculty of Horticulture and Landscape Architecture, Warsaw University of Life Sciences - SGGW, Nowoursynowska 159, Warsaw, 02-606, Poland

New sequencing technologies make it possible to achieve genome sequence fast and cheap. Since the assembly step of such next generation reads is still not well standardized it is the most cumbersome part of sequencing projects. We present here approaches undertaken on the way to the draft assembly the cucumber (Cucumis sativus L. cv. Borszczagowski) genome done with the use of 8x unpaired and 4x paired (3 Kbp) pyrosequenced 454 XLR Titanium reads, together BAC library ends fragments (12,7x physical coverage).

Two different assembly approaches, namely Celera and Celera/Arachne were finally used. The Celera result was 15'196 contigs of 197 Mbp length and N50 27'086 bp and 4'157 scaffolds of 224 Mbp length and N50 2'324 Kbp. In the Celera/Arachne method contigs after prior pyrosequencing reads assembly (in the form of \sim 800 nt reads) were used together with the STCs as the input to Arachne assembler. Number of contigs were 15'764 of the length 193 Mbp with N50 23'280 bp. Number of supercontigs were 12'438 with the coverage of 323 Mbp and N50 323'092 bp.

Correctness of the assemblies were proven after mapping 95,56% of 63'035 cucumber unigenes with the average identity 97,81%. Additionally 6 BAC/Fosmid cucumber sequences (totally 372'277 bp) had identity of 97,61% to the assembled genome. Average reads coverage of the genome were 14,20x and 98% of the assembled genome had reads coverage higher than 3x.

Taking into consideration coverage used in already reported cucumber assemblies (32x by Miller J. et al., 2009 and 72x by Huang S., et al., 2009), as well as no differences in the quality and the best overall quantity statistics of contigs/scaffold lengths, the Celera approach used in this project should be considered the most optimal one to get omics-ready quality draft sequence of highly repeated eukaryotic genome.

Resequencing or de novo sequencing? Targeted sequence finishing for the proteomic analysis of glycosylhydrolases from Caldicellulosiruptor spp.

<u>David E. Graham</u>, Karen Walston Davenport, James Elkins, Richard Giannone, Lynne Goodwin, Loren Hauser, Robert Hettich and Miriam Land, and Adriane Lochner

Oak Ridge National Laboratory, Oak Ridge, TN

The Caldicellulosiruptor genus includes thermophilic, anaerobic bacteria that efficiently hydrolyze cellulose and hemicellulose polymers. These fermentative microbes could become valuable catalysts for biofuel production through consolidated bioprocessing, or as sources of thermoactive glycosylhydrolase enzymes. With complete genome sequences for Caldicellulosiruptor bescii (formerly "Anaerocellum thermophilum" DSM 6725) and C. saccharolyticus, this BESC project will produce complete genome sequences of six more species for comparative analysis, proteomics and transcriptional profiling. These genome sequences vary from 78 to 98% pairwise average nucleotide identity, with broad gene synteny. However, large scale chromosomal rearrangements, and significant variability in a crucial island encoding most cellulose-degrading enzymes complicated genome assembly. Highly repetitive carbohydrate binding modules in these genes have hindered the assembly of these regions. We show the utility of comparative analysis for targeting finishing efforts to interpret quantitative proteomic data from C. obsidiansis. The combined data illustrate a new bacterial strategy for lignocellulose degradation that incorporates aspects of both fungal and clostridial cellulases.

"The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231."

Use of optical mapping in bacterial genome finishing

Dibyendu Kumar and William G Farmerie

Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL-32611

The cost-efficiency of modern DNA sequencing technology, such as the Roche 454 GS-FLX, allows individual investigators to undertake bacterial genome projects that were not affordable only a few years ago. Our core laboratory has several ongoing bacterial genome projects presenting a variety of challenges to genome assembly and closure. Several factors contribute to these challenges; including sequence repeats versus read length, intrinsic sequencing errors, and dynamic genome rearrangements. Together these factors complicate genome closure when using shotgun DNA sequencing alone. The genome finisher may experience difficulty validating their assembly in the absence of a physical map. To address this problem, we adopted whole-genome optical mapping as a tool to validate bacterial genome assemblies. OpGen, Inc. (Gaithersburg, Maryland) prepared the optical maps used in this project.

Briefly, an optical map is a complete genome restriction map deduced from a number of partial restriction maps. Optical maps are generated by spreading carefully extracted genomic DNA onto a treated glass surface containing many narrow channels, followed by digestion *in situ* with restriction enzymes. About 50–100 contiguous restriction fragments with sizes approaching up to one-third of the whole genome are selected and optically measured. The overlapping partial optical contigs are combined by alignment software to produce a contiguous whole genome restriction map. The contiguous optical map can be aligned and compared with the *in silico* restriction map determined for the partially complete whole-genome assembly. We successfully used optical mapping for guiding the closure of four closely related bacterial genomes. The optical map allowed us to identify assembly errors not possible using shotgun DNA sequencing data alone. Thus, we conclude that, in order to ensure the accuracy of a finished bacterial genome, optical mapping is an important tool to validate *de-novo* assemblies generated by next-generation DNA sequencing.

Bacterial biodiversity and function in a cold desert ecosystem

<u>Cristina Takacs-Vesbach</u>¹, David Van Horn¹, Lydia Zeglin^{1,2}, Shannon FitzPatrick¹, Michael Gooseff³, and John Barrett⁴

For many decades the soils of the McMurdo Dry Valleys, Antarctica were thought to be essentially sterile. We now know that this is an ecosystem that is dominated by microorganisms, however, early cultivation efforts largely failed to detect the biodiversity of the region's poorly weathered, low organic carbon soils. Initial surveys of microbial diversity using 16S rRNA gene sequencing has revealed a surprising bacterial richness, including representatives from at least ten different phyla, including a high proportion of unique and rare sequences. Yet, a thorough and exhaustive survey of microbial diversity has not been conducted and little is known about the function of the detected microorganisms. Furthermore, given the low in situ microbial activity and decomposition rates, the question of whether this richness is illusory has been raised. We are conducting an exhaustive survey of the microbial richness, function, and activity of soil bacteria across gradients of moisture and salinity using pyrosequencing of 16S rRNA bacterial tag-encoded FLX amplicons (bTEFAP) and environmental DNA (metagenomics). These data will be used to identify the extent of active bacteria in dry valley soils and elucidate potential microbial function with the ultimate goal of understanding the role of bacteria in cold arid soils.

¹Department of Biology University of New Mexico Albuquerque, NM 87131

²Department of Crop and Soil Science Oregon State University Corvallis, OR 97331

³Department of Biological Sciences Virginia Tech Blacksburg, VA 24061

⁴Department of Civil & Environmental Engineering Pennsylvania State University University Park, PA 16802

Investigation of Virulence and Possible Expanded Host Range in *Rickettsia* species by High-Throughput Sequencing

K. A. BISHOP-LILLY ¹, H. GE¹, P. MISHRA¹, M. PATEL¹, K. M. WILLNER ¹, A. BUTANI ¹, S. DORSEY ¹, M. GEORGE ¹, T. D. READ ², T. N. BRAHMBHATT ¹, A. MATECZUN ¹, S. SOZHAMANNAN ¹, A. L. RICHARDS ¹

¹Naval Med. Res. Ctr., Silver Spring, MD, ²Naval Med. Res. Ctr., Silver Spring, MD. Current address: Emory GRA Genomics Core, Emory Univ. Sch. of Med., Atlanta, GA.

Rickettsiae are Gram negative, obligate intracellular bacteria that cause a variety of human diseases transmitted by arthropods. *Rickettsia prowazekii* is a NIAID Category B priority pathogen for biodefense research and has a history of incapacitating military troops. We have undertaken whole genome sequencing of various R. prowazekii isolates using 454 high-throughput sequencing in order to understand certain differences between the isolates, such as host preference and virulence. Specifically we performed whole genome sequencing (WGS) on several R. prowazekii isolates, including an avirulent type strain, a virulent strain, and a strain isolated from a flying squirrel (a potential example of expanded host range). Using these draft sequences, we have identified putative differences among the three genomes that may account for phenotypic differences in terms of reservoirs and virulence. Additionally, for two of the strains, we were able to close the genomes to one contig without performing any additional wet lab experiments. Subsequently, we performed metagenomic sequencing of arthropods to detect and characterize rickettsial agents in their vectors. To assess metagenomic sequencing for identification of Rickettsiae in ticks, total DNA was extracted from ticks and sequenced. For one tick, of the reads that could be classified by MG-RAST 1,311 were classified as being derived from the Rickettsiales. Interestingly, these reads had their best hits to a variety of rickettsial species and 2 different Wolbachia species. Species with the most hits were R. felis and R. bellii, although there were notable differences in the overall coverage of these genomes. These results appear consistent with previous studies that have shown presence of R. felis and R. bellii. These results highlight the potential of nextgen sequencing technologies for application to not only rapid characterization of virulence genes and other factors by WGS but also for vector surveillance.

Hybridization Assisted Nanopore Sequencing

<u>John S. Oliver</u>, Heidi Geiser, Peter lanakiev, Peter Goldstein, Maryam Jouzi, Peter H. Goldstein

NABsys, Inc., Providence, RI 02903

NABsys Inc. is developing a nanopore based DNA sequencing instrument. In the past decade it has been shown that nanopores are sensitive resistive pulse detectors capable of discriminating between individual DNA molecules based on length, state of folding, or duplex versus single stranded forms.

NABsys' technology, termed Hybridization Assisted Nanopore Sequencing (HANS) relaxes the stringent requirement that the nanopore have single base resolution. The nanopore is used to detect hybridization of probes to target DNA and to give relative positional information. The positional accuracy required for assembly of the sequence is reduced by orders of magnitude with respect to direct nanopore sequencing.

We will discuss progress towards implementing HANS using a procedure that allows us to detect probes bound to single-stranded template DNA with high signal-to-noise and high resolution. We will discuss the sample preparation procedure, results of translocation experiments and the ramifications for correct assembly and reconstruction of sequence information.

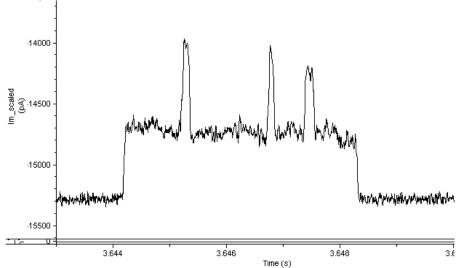


Figure 1: Three probes bound to M13 at positions 1600 bases, 3850 bases, and 4850 bases from the end of the template strand.

The Genome Sequencing Center at NCGR

Ryan W. Kim, Jimmy E. Woodward, Peter B. Ngam and Faye D. Schilkey

National Center for Genome Resources, Santa Fe, NM

National Center for Genome Resources (NCGR) is a non-profit research institute whose mission is to improve human health and nutrition through genome sequencing and analysis. The Genome Sequencing Center at NCGR offers next generation sequencing, genotyping, and analysis services, currently housing eight Illumina Genome Analyzers IIx for sequencing and Illumina's BeadExpress and iScan for genotyping. Sequencing services include whole genome and transcriptome shotgun sequencing, ChIP, small RNA, whole or targeted exome, and methylome sequencing. This summer, NCGR is one of only 10 sites including the Broad institute of MIT and Harvard, CSHL, JGI, and Wash U who are receiving the third generation Pacific Biosystem single molecule sequencing instrument. The PacBio system boasts very long nucleotide read lengths that provide fidelity especially helpful for de novo sequencing and resequencing projects.

NCGR has a longstanding reputation for developing effective bioinformatics tools for scientific research evident by two recent awards: 2009 Bio-IT World Best Practices in Basic Research in Schizophrenia and 2009 Laureate by Computerworld for Alpheus® a web-based analysis pipeline for variant and expression detection of next generation sequencing data.

Leveraging NCGR's own research, sequencing and analysis experience, the NCGR Genome Sequencing Center provides an unprecedented knowledgebase and facility with approximately 100 customers/collaborators to date from around the globe. Please contact Faye D. Schilkey at seq@ncgr.org to find out how NCGR can accelerate your research.

FUNCTIONAL ANNOTATION of Clostridium thermocellum GENOME Applying 454 sequencing to phage display selections

<u>Sara D'Angelo^{1*}</u>, Flavio Mignone ^{2*}, Nileena Velappan ¹, Tiziano Gaiotto ¹, Andrew Bradbury ¹

Los Alamos National Laboratory, Los Alamos, NM ^{1;} University of Milan, Milan, Italy ² *Equal contributors

454 sequencing was applied to a Phage display selection of genomic DNA from *Clostridium thermocellum*. This is a thermophilic and anaerobic gram-positive bacterium with the highest rate of cellulose utilization of any known bacterium [1], whose genome was recently resequenced (US DOE JGI, 2007) and partially annotated. DNA encoding protein domains were first filtered from the genomic DNA and then displayed on phage and selected on microcrystalline cellulose in order to identify known and new cellulose binding domains (CBD).

The limit in Phage display selection is in the number of clones that can be screened by means of functional assays; we demonstrated that the application of a bioinformatic algorithm to 454 sequencing outputs allowed us analyze the whole phage display selection and identify the most enriched clones as strong CBDs.

We have shown in this work, and in the past [2], the power that the application of 454 deep sequencing can bring to library analysis, and, in particular, as it relates to the analysis of protein/substrate and protein/protein interactions. Moreover, beyond the identification of specific proteins involved in interactions, the fragmented nature of the library allows information to be obtained at the domain level.

The first clear advantage of this approach is the reduction of the screening effort. Furthermore, it appears feasible to use 454 sequencing applied to selections carried out on a wide variety of targets to carry out experimental functional annotation on the basis of phage selection outputs.

References:

- [1] Gold et al, Journal of Bacteriology, 2007
- [2] Di Niro et al, NAR, 2010

Experimental gene annotation by deep sequencing: Identifying cellulose binding domains in *Clostridium thermocellum*

N. Velappan, S. D'Angelo, T. Gaiotto, F. Mignone and A.R.M. Bradbury

Los Alamos National Laboratory, Los Alamos, NM 87545 University of Milan, Milan, Italy

Many proteins have binding as part of their function, which is usually encoded by separate domains. In this project we have developed the experimental and computational methodology to identify specific binding protein domains from a genome, using cellulose binding domains (CBD) from the Clostridium thermocellum genome as a model system. The first step is to "filter" the genome for encoded protein domains. This is carried out using open reading frame (ORF) filtration vectors in which DNA fragments are cloned between a leader sequence and the □-lactamase gene. If the cloned DNA fragment encodes an open reading frame and all three components are correctly translated and functional, the bacteria will survive on ampicillin. Higher concentrations of ampicillin will select better folded protein domains. Three series of vectors with different leaders (pelB, TAT, SRP) were constructed to isolate domains with different folding properties. After open reading frame filtration, isolated fragments were cloned into corresponding phage display vectors using p3 in M13 as a display protein. After selection on cellulose, ELISA screening and 454 based deep sequencing were used to identify cellulose binders. We found four strong CBDs in C.thermocellum. Two of them (CBD of Cthe 3077 and Cthe 2972) were identified in random screening and are the most enriched genes in 454 analyses of pelB and SRP libraries respectively. In addition 454 analyses facilitated the identification of two new CBDs. Currently, Cthe-1068 is described as hypothetical protein and Cthe 2159 is designated as cytidine deaminase. This work indicates that experimental gene annotation can be carried out by a combination of phage display and 454 sequencing.

DNA Production Sequencing at the Joint Genome Institute: A Multi-Process Operating System

Miranda Harmon-Smith1, Simon Roberts2

1. Lawrence Livermore National Laboratory, 2. Lawrence Berkeley National Laboratory US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA

The U.S. Department of Energy (DOE) Joint Genome Institute's (JGI) Production Sequencing group is committed to the generation of high-quality genomic DNA sequence to support the mission areas of renewable energy generation, global carbon management, and environmental characterization and clean-up. In 1999, the Joint Genome Institute opened its doors to the world of production sequencing with its participation in the Human Genome Project. This project set in motion a major change in how DNA sequencing was performed. A mostly manual process was turned into a fully automated production line at the JGI, allowing for an unprecedented amount of DNA sequence to be provided to the scientific community at large. The advances made in technology during and after the Human Genome Project, and the continuous process improvements made to our workflow design, helped to push forward the ability to obtain the equivalent of the human genome (3 billion bases) in only one month time. From 1999 until 2009, the JGI performed this work largely through a process known as Sanger sequencing using capillary sequence technology. Today we run Sanger sequencing at a greatly reduced capacity, along with two new platforms which utilize pyro-sequencing and sequencing by synthesis technologies. The Roche Titanium Genome Sequencer and Illumina Genome Analyzer pipelines have allowed us to increase our sequencing capacity to new levels, 1 trillion bases in a month! What took us 1 month to sequence using capillary technology can now be accomplished in one week. With these continuous advances and changes in our production pipeline, come new challenges. Our production group is now focused on how to schedule and operate a 3 platform multiprocess operating system, on a daily basis, while maintaining the safety of our staff and the quality of our sequence data. Presented here will be the Sanger, Roche, and Illumina pipelines and our current strategies for scheduling and operation, along with current optimization projects underway. This work is a joint effort among the Process Optimization, Sequencing Technologies, Instrumentation & Industrial Engineering, and the core Sanger, Roche and Illumina Production group.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

Automation of Finishing Project Startup Tasks

Y. Xu, L. Meincke, R. Arnaudville, E. Davydenko, A. Seirp, R. Tapia, C. Han, <u>J.</u> Challacombe.

Genome Science Group (B-6), Los Alamos National Laboratory. Los Alamos, NM 87544

New sequencing technologies brought many challenges to finishing: short reads, massive amounts of data, new assemblers, and data quality challenges. At LANL, these were addressed with a series of informatics tools: new assemblers, incorporation of JGI-developed code to post-process assembly data and support phrap assemblies, and new tools to examine assembly data, QC, adjust for quality, and propose finishing reactions. These steps were fairly routine, but also time and resource intensive, for incoming finishing projects. When FinishingStartup was initiated in June 2009, the goal was to automate those sequential startup tasks to allow: 1) Finishers to focus on the intellectual challenges of finishing, rather than routine tasks, 2) optimization of computer resources, by delegating cpu- and memory-intensive tasks to specialized servers, and 3) increase the size and decrease the number of contigs in projects delivered to the LANL Finishing Team. The system was put into place in August of 2009. The process has been refined since then, and is continually being refined in order to provide the best possible inputs into the finishing process.

The JGI-LANL Informatics system built upon the existing infrastructure, using the Pipes and Filter and the Broker Architectural Patterns. A new Finishing Startup Manager was developed along with a Monitor to track each FinishingStartup task as it moves through the automated steps of project initiation. The Finishing Startup Manager relies on existing servers, queues and other mangers in a complex web of services, which are linked together into one process. The result is the same, mostly better and consistent output, delivered automatically, with time determined by process speed, without requiring human input. When the project is delivered, the human interaction begins. The finisher will then examine the current state to determine which of the suggested reactions should be submitted for further improvement of the genome assembly.

Ecological and Evolutionary Relevance of the Genomes of Thermophilic Fungi

Amy J. Powell,^{2,1} Joslyn Bustamante,¹ Eric Ackerman,² Bryce Ricken,² Randy Berka,³Adrian Tsang⁴ and **Donald O. Natvig**¹*(dnatvig@unm.edu)

¹Department of Biology, University of New Mexico, Albuquerque; ²Sandia National Laboratories, Albuquerque, NM; ³Novozymes, Inc., Davis, CA; ⁴Center for Functional and Structural Genomics, and Department of Biology, Concordia University, Montreal, Canada

Although thermophilic fungi have received substantial attention for potential use in industrial applications, important questions remain regarding their distributions, ecology, life cycles and diversity. If these organisms are to be exploited fully in industry, including the production of biofuels, it will be important to better characterize aspects of their fundamental biology. We are examining the ecology and evolution of major groups of thermophilic fungi, combining information from new genome sequences with field and laboratory studies of isolates obtained from arid ecosystems. Within the Ascomycota, thermophily has arisen independently in at least two groups, the Chaetomiaceae and the Eurotiales. The genomes of three closely-related members of the Chaetiomiaceae have been sequenced, and at least eight well-curated genomes are available for mesophilic and thermolerant taxa within the Eurotiales. sequenced members of the Chaetomiaceae, Thielavia terrestris and Sporotrichum thermophile, are thermophiles, while the third, Chaetomium globosum, is not. Analysis of these genomes underscores key unresolved issues within the family. One is that taxonomic and nomenclatural problems are rampant among thermophilic species. For example, in many cases a single genus name has been applied to species from distantly related groups. This includes Sporotrichum, a name that has been used for diverse filamentous fungi and is perhaps properly applied to particular anamorphic species in the Basidiomycota. Moreover, it would appear that the genus *Chaetomium* as currently viewed is paraphyletic. Aside from engendering confusion about naming, the poor understanding of phylogenetic relationships within the Chaetomiaceae makes it impossible to evaluate where thermophily has been gained or lost in the family, which in turn hampers efforts to probe the molecular basis of thermophily. Our field studies of litter, biological crusts, rhizosphere soils and herbivore dung from the Sevilleta Long-Term Ecological Research Site in central New Mexico reveal a large diversity of thermophilic fungi from both the Chaetomiaceae and the Eurotiales. conclusion based on analyses of ribosomal ITS and other sequences is that species diversity in both major groups of thermophilic fungi has been underestimated.

Development of High Throughput Processes for Constructing 454 Titanium and Illumina Libraries

<u>Shweta Deshpande1</u>, Chris Hack1, Eric Tang1, Stephanie Malfatti2, Aren Ewing1, Susan Lucas2, Jan-Fang Cheng1

1. Lawrence Berkeley National Laboratory, 2. Lawrence Livermore National Laboratory US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA

We have developed two processes with the Biomek FX robot to construct 454 titanium and Illumina libraries in order to meet the increasing library demands. All modifications in the library construction steps were made to enable the adaptation of the entire processes to work with the 96-well plate format. The key modifications include the shearing of DNA with Covaris E210 and the enzymatic reaction cleaning and fragment size selection with SPRI beads and magnetic plate holders. The construction of 96 Titanium libraries takes about 8 hours from sheared DNA to ssDNA recovery. The processing of 96 Illumina libraries takes less time than that of the Titanium library process. Although both processes still require manual transfer of plates from robot to other work stations such as thermocyclers, these robotic processes represent about 12- to 24-folds increase of library capacity comparing to the manual processes. To enable the sequencing of many libraries in parallel, we have also developed sets of molecular

barcodes for both library types. The requirements for the 454 library barcodes include 10 bases, 40-60% GC, no consecutive same base, and no less than 3 bases difference between barcodes. We have used 96 of the resulted 270 barcodes to construct libraries and pool to test the ability of accurately assigning reads to the right samples. When allowing 1 base error occurred in the 10 base barcodes, we could assign 99.6% of the total reads and 100% of them were uniquely assigned. As for the Illumina barcodes, the requirements include 4 bases, balanced GC, and at least 2 bases difference between barcodes. We have begun to assess the ability to assign reads after pooling different number of libraries. We will discuss the progress and the challenges of these scale-up processes.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 LLNL-ABS-420646

LBNL-2902E Abs

Large Gap Size Paired-end Library Construction for Second Generation Sequencing

<u>Ze Peng,</u> Matthew Hamilton, Jeff Froula, Aren Ewing, Brian Foster, and Jan-Fang Cheng

US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598 USA

Fosmid or BAC end sequencing plays an important role in de novo assembly of large genomes like fungi and plants. However construction and Sanger seguencing of fosmid or BAC libraries are laborious and costly. The current 454 Paired-End (PE) Library and Illumina Jumping Library construction protocols are limited with the gap sizes of approximately 20 kb and 8 kb, respectively. In the attempt to understand the limitations of constructing PE libraries with greater than 30Kb gaps, we have purified 18, 28, 45, and 65Kb sheared DNA fragments from yeast and circularized the ends using the Cre-loxP approach described in the 454 PE Library protocol. With the increasing fragment sizes, we found a general trend of decreasing library quality in several areas. First, redundant reads and reads containing multiple loxP linkers increase when the average fragment size increases. Second, the contamination of short distance pairs (<10Kb) increases as the fragment size increases. Third, chimeric rate increases with the increasing fragment sizes. We have modified several steps to improve the quality of the long span PE libraries. The modification includes (1) the use of special PFGE program to reduce small fragment contamination; (2) the increase of DNA samples in the circularization step and prior to the PCR to reduce redundant reads; and (3) the decrease of fragment size in the double SPRI size selection to get a higher frequency of LoxP linker containing reads. With these modifications we have generated large gap size PE libraries with a much better quality. We are also testing the construction of large gap size PE libraries with the Illumina adaptors. The result of the test will be discussed.

JCVI's Viral Finishing Pipeline: Incorporating Next Gen Sequencing Data and Automation

E. Hine, N. Fedorova, J. Sitz, D. Katzel, T. Stockwell, L. Overton, B. Szczypinski, <u>J. Hostetler</u>, R. Halpin, D. Spiro

The J Craig Venter Institute, Rockville, MD, U.S.A.

The viral finishing pipeline at JCVI traditionally employed amplicon-based PCR Sanger sequencing and increasingly automated data processing steps to complete over 4500 Influenza A and B genomes as part of the Influenza Genome Project supported by the NIAID Genomic Sequencing Center for Infectious Disease (GSCID). Recently, over 1000 H1N1pdm samples were closed using Sanger sequencing and the latest automated processing software, detailed below. JCVI is also transitioning to Next Gen sequencing technologies, specifically for use with diverse avian influenza samples.

Currently, two automated software suites are integrated into a Sanger pipeline designed for H1N1pdm Influenza samples. The first, **Vapor**, is an assembly software suite that loads, assembles, and validates Influenza and other viral samples. The second is **autoTasker**, which examines samples for low coverage and low quality using **contigChecker**, and then automatically assigns PCR tasks needed to finish samples. Our ultimate goal is to combine these programs into a single, integrated software suite for rapid and efficient complete genome sequencing of Influenza and other viruses. The combination of highly optimized Sanger sequencing and an automated software suite allows for large volumes of sample processing with limited manual interaction.

The above software development is also integrated into a new Avian Influenza sequencing pipeline that utilizes SISPA-generated genomic libraries and Next Gen sequencing, including 454 and Illumina technologies. The new sequencing technologies allowed us to more efficiently and completely sequence traditionally challenging avian influenza genomes. The software is also incorporated into the finishing and analysis of these influenza genomes, allowing us to maintain an automated sample processing pipeline that requires little RA interaction. In addition, finishing work is now performed using the widely available **Consed** editor or could be done using other, flat-file based editors.

All of these new developments have significantly increased production, effectively decreasing costs, manual labor, and sample completion time.

Rapid Finishing of Yersinia pestis KIM D27 at JCVI

D. Radune, J. Hostetler, M. Kim, J. Varga, W. Nierman

The J Craig Venter Institute, Rockville, MD, U.S.A.

Yersinia pestis KIM D27 strain is an avirulent strain of Yersinia pestis, the etiologic agent of plague. The KIM D27 strain, a derivative of the fully virulent strain KIM 10, contains a 100kbp deletion that encodes several genes including genes for iron acquisition. The KIM D27 strain sequenced at JCVI was isolated from a lab where a researcher died from an accidental infection by the D27 strain. JCVI was tasked to rapidly sequence and analyze the KIM D27 strain and to investigate any differences between KIM D27 and KIM 10 that may have contributed to the sudden virulence and death.

The strain was sequenced with a full plate 454 Titanium 8 kb paired-end run and a single lane of 100 base paired-end Illumina. The average coverage of the chromosome was 400x. In addition, Sanger sequences were generated to fill the gaps and produce a final finished molecule as well as to confirm differences between the D27 strain and the KIM 10 reference (NCBI gi|22002119|gb|AE009952.1|). A number of genome assemblers were first used with the 454 only data with variable results. Most of the assemblies were very fragmented with a high degree of discrepant bases relative to the KIM 10. Illumina sequencing was added to correct consensus errors and improve the overall assembly quality. A total of 57 chromosomal SNP/INDELs and one plasmid SNP were identified relative to the KIM 10 reference and these were verified by sequencing of PCR products. In addition, several repeat areas were verified to be different in length and repeat number compare to the KIM 10 sequence.

In order to verify that the detected repeat and SNP differences new in the D27 strain, the *Y. pestis* KIM10+ DNA was obtained from ATCC and all SNP and repeat PCRs were run on this DNA. This poster will highlight the JCVI finishing process of *Y. pestis* KIM D27, the results of the comparative analysis between KIM D27 and KIM 10 strains and some challenges that finishing group faced during this process. It will also review the consequence of the 454 homopolymer induced sequencing errors in performing this analysis.

Finishing of Human Microbiome Project (HMP) Reference Strains at JCVI

N. Fedorova, D. Radune, M. Kim, B. Szczypinski, M. Torralba, J. Hostetler

The J Craig Venter Institute, Rockville, MD, U.S.A.

The <u>Human Microbiome Project</u> (HMP), funded by NIH, studies the microbial communities living in and on the human body in order to understand their role in human health and disease. A major objective of the HMP is to define a core set of microbial species that are associated with the human body known as the "core microbiome". JCVI was selected as one of four sequencing centers to generate a microbial genome reference set of at least 1000 genomes and characterize the microbial communities from multiple body sites. In 2009, JCVI's Genome Finishing and Analysis group improved 11 microbial strains, which represent 22% of the 50 strains sequenced during the project's Jumpstart phase. Seven of these genomes were fully finished. The Jumpstart component ended December 31st, 2009, and the project has fully entered the large scale phase. JCVI's Genome Finishing and Analysis group is scheduled to improve up to 30 additional genomes to *Improve High-Quality Draft*, *Noncontiguous Finished* or *Finished* levels.

Sequencing of HMP genomes at JCVI primarily uses 454 fragment or paired-end sequencing technology. Early genomes contained low coverage Sanger sequencing. While in the past, JCVI's finishing strategy relied heavily on database-dependent inhouse software, it is now moving towards widely-used tools and flat files. HMP Jumpstart genome, Neisseria flavescens SK114, was finished exclusively using Consed. Finishing HMP genomes presents many challenges mainly because they are often phylogenetically novel. Culturing samples and obtaining DNA can be very difficult. At least two HMP genomes scheduled for finishing were discovered to be a mixture of two strains. This poster will highlight the finishing process of HMP genomes at JCVI as well as various challenges that the finishing group encountered.

Robotic Enrichment Processing of Roche 454 Titanium Emulsion PCR at the DOE Joint Genome Institute

<u>Matthew Hamilton</u>¹, Steven Wilson¹, Don Miller², Kecia Duffy-Wei¹, Nancy Hammon¹, Susan Lucas², Martin Pollard¹, Jan-Fang Cheng¹

1. Lawrence Berkeley National Laboratory, 2. Lawrence Livermore National Laboratory

US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA

Enrichment of emulsion PCR product is the most laborious and pipette-intensive step in the 454 Titanium process, posing the biggest obstacle for production-oriented scale up. The Joint Genome Institute has developed a pair of custom-made robots based on the Microlab Star liquid handling deck manufactured by Hamilton to mediate the complexity and ergonomic demands of the 454 enrichment process. The robot includes a custom built centrifuge, magnetic deck positions, as well as heating and cooling elements. At present processing eight emulsion cup samples in a single 2.5 hour run, these robots are capable of processing up to 24 emulsion cup samples. Sample emulsions are broken using the standard 454 breaking process and transferred from a pair of 50ml conical tubes to a single 2ml tube and loaded on the robot. The robot performs the 454 enrichment protocol and produces beads in 2ml tubes ready for counting. The robot follows the Roche 454 enrichment protocol with slight exceptions to the manner in which it resuspends beads via pipette mixing rather than vortexing and a set number of null bead removal washes. The robotic process is broken down in similar discrete steps: First Melt and Neutralization, Enrichment Primer Annealing, Enrichment Bead Incubation, Null Bead Removal, Second Melt and Neutralization and Sequencing Primer Annealing. Data regarding machine to machine variability and validation studies will be discussed.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

Covaris AFA-based DNA Fragmentation: A Critical First Step for the Optimization of Next Gen Sequencing

Jim Laugharn, Guillaume Durin, and Edwin Rudd

Covaris, Inc., 14 Gill Street, Unit H, Woburn, MA 01801 USA

The controlled generation of random fragments from genomic DNA, with high recovery and high quality, is the most critical sample preparation step required by all the currently available next generation sequencing platforms. As the requirements for improved sensitivity and application depth of these platforms increase, high demands are also placed on the standardization of the sample fragmentation process to be easily transferable between laboratories (intra- and inter-lab) and scalable to enable both automation and higher throughput requirements. To be successful in the next-gen sequencing market, it requires utilizing an efficient, easy-to-use, high throughput, and highly reproducible DNA shearing technology. Standardizing on high accuracy and high precision will become a fundamental prerequisite of all nucleic acid fragmenting sample preparation processes, as genomic sequencing technologies continue to develop; particularly in clinical trials and clinical applications.

Although nebulization (wind shear), enzymatic digestions, hydrodynamic shearing, and low frequency sonicators have been used to shear DNA, they all have significant intrinsic disadvantages which make them the weak points in next generation sequencing workflows. These limitations and disadvantages include; thermal and sequence specific biased fragmentation, thermal degradation, valuable sample loss, automation incompatibility, and user-dependent reproducibility issues. In stark contrast, the closed vessel, non-contact, isothermal processing, and easily automatable Covaris Adaptive Focused Acoustics™ (AFA) process stands out as the most widely used and preferred method for easily and reproducibly shearing DNA.

We will show how AFA enables standardization of a sample preparation process and work flow. We will present data which demonstrates the reproducibility of our validated protocols to efficiently fragment DNA in the 100bp to 5kb size range, and we will discuss the use of our technology in chromatin shearing for ChIP-chip, ChIP-Seq applications. Performance data from new products will be presented on a designed—for-automation, pre-assembled 96 microTUBE plate and the parallel LE220 instrument (automation-ready) for high-throughput laboratories.

Gold Standard: Finishing Using Next Generation Technologies

Andries J. Van Tonder

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA, UK

A pipeline for finishing genomes to published criteria has been designed using *Staphylococcus lugdunensis*. This pipeline is scalable and applicable to a variety of data sets consisting of next generation sequencing technologies.

S. lugdunensis is a Gram-positive bacterium which occurs as a commensal on human skin. The 5Mb genome was sequenced using both the Illumina GAII platform to a depth of 630x coverage and the 454/Roche Titanium FLX3 platform to a depth of 16x coverage. Illumina data was assembled using Velvet before IMAGE (Iterative Mapping and Assembly for Gap Elimination; Tsai In press) improved the assembly by iterative gap closure. Velvet consensus sequences were combined with 454 data to create an assembly using Newbler. The Ace file generated by Newbler could then be converted to be viewable in a sequence editor such as consed or gap4. ABACAS (Algorithmic Based Automatic Contiguation of Assembled Shotgun Sequence; Assefa 2009), along with the 454scaffold.fna file from Newbler, identified the correct contig order. This allowed manual finishing, including the ordering of reactions such as PCRs, to be undertaken. Once the genome was contiguous scripts such as iCORN (Iterative Correction of Reference Nucleotides; Otto Unpublished), which provided sequence correction by mapping Illumina reads back to the assembly, were used to improve the quality of the data.

A set of criteria was applied to ensure that the genome was finished to Gold Standard, as detailed by Chain et al. (2009). This requires every position in the sequence be covered by at least two technologies and the error rate of the consensus needs to be less than 1 error per 100,000 base pairs. The assembly was checked to ensure that it was free of misassemblies and that repeat regions had been ordered and correctly assembled.

The described pipeline will be presented detailing the work flow and methodologies.

FF0101, #2

Variant Validation and Screening Methods at the Genome Center at Washington University School of Medicine

<u>Bob Fulton</u>, Li Ding, Vincent Magrini, Michael D. McLellan, Daniel Koboldt, Heather Schmidt, Michelle O'Laughlin, Rachel M. Abbott, Timothy J. Ley, Elaine R. Mardis, and Richard K. Wilson.

The Genome Center at Washington University, 4444 Forest Park Blvd., Saint Louis, MO 63108

The emergence of next generation sequencing technologies has resulted in an explosion of data production capabilities and the capability to study entire human genomes on a daily basis. With this massive sequencing capacity expansion, the ability to call, validate, and screen additional samples for putative variants is essential. Although variant detection methods are ever improving, with greater sensitivity and specificity, the investment in generating the primary data, as well as the need to understand each genome's exact mutation spectrum, means that validation of putative variants is still a necessary component. The Genome Center at Washington University has developed a robust, multi-platform, cost-effective validation platform, with the capability of validating variants at a comparable pace as the raw data production capabilities. The validation process is capable of leveraging Illumina, Roche 454, and ABI 3730 sequencing instruments. Template targeting is provided by either PCR or hybrid-capture. Both platform selection and template generation methods are determined based on scale, timeline, and data output requirements. The composition of the variant lists can include single nucleotide variants (SNV), small (< 20-30bp) insertion or deletion events, as well as structural variations including large (> 20-30bp), insertions or deletions, translocations, and inversions. Validation can be performed on single samples, panels of individuals that maintain individual genotypes, or pools of individuals providing pooled genotypes. The use of sample tags or DNA barcodes are utilized when possible, to maintain as much sample-to-genotype specificity as possible. Pooling strategies and sample barcodes are particularly useful when working with sample pools or panels, although they also can be applied for small sample sets such as tumor and normal samples. These barcodes enable more efficient use of the large sequence throughput per run provided by the next generation sequencing instruments. In addition to current technologies, new methods and instrumentation are in development with new validation and screening procedures coming soon. Without these developments and continued confirmation of the detection methods, improvements and significant findings will not likely be produced as quickly, accurately, or effectively. This presentation will highlight many of the current methods and ongoing work to improve these processes.

Separation of Propionibacterium acnes genomes HMP110/SK182

<u>Sergey Koren</u> (1), Scott Durkin (1), Marcus Gillis (1), Jessica Hostetler (1), Jamison McCorrison (1), Jason Miller (1), Guillermo Perez-Perez (2), Diana Radune (1), Julian Tatz (2), Manolito Torralba (1), Martin Blaser (2), Granger Sutton (1), Karen Nelson (1)

(1) The J. Craig Venter Institute, Rockville, MD, (2) New York University Langone Medical Center, New York, NY

Propionibacterium acnes is a Gram-positive bacterium abundant in the human cutaneous microbiome. The genome sequence of the KPA171202 strain was published in 2004. Four other strains, isolated at NYU Medical Center and sequenced at JCVI, had shown consistent genome sizes of about 2.5 Mbp. A fifth strain, SK182, underwent 454 Titanium unpaired shotgun sequencing expected to generate 86X coverage. The assembled result spanned 4.2 Mbp unexpectedly. 16S sequencing had revealed no contamination in the DNA. Alignment of reads to the KPA171202 reference revealed two populations with about 99% and 89% identity, respectively. Based on a hypothesis that two similar strains were present in the sequenced sample, the reads were partitioned by read similarity to the reference and contig depth in the initial assembly. Each partition was assembled separately and aligned to the KPA171202 reference. There were no rearrangements, few coverage gaps, and consistent levels of coverage within both assemblies. Average coverage was 55X and 30X, respectively. These results indicate that a partition-and-assemble method successfully recovered two complete genomes from a sample containing two highly similar members of the same species. These results have positive implications for future analysis of metagenomics samples and uncultured samples, both of which are essential components of the Human Microbiome Project.

Metagenomic Assembly Strategies

<u>Sergey Koren</u> (1), Doug Rusch (1), Jason Miller (1), Granger Sutton (1), Mihai Pop (2) (1) The J. Craig Venter Institute, Rockville, MD (2) The University of Maryland Center for Bioinformatics and Computational Biology, College Park, MD

Metagenomic assembly (the simultaneous assembly of DNA from all organisms present in an environment) enlarges the substrate for analysis but undermines the assumptions made by current assembly software. Metagenomic datasets present variations in coverage independent of the repetitiveness of a region within an organism. Large-scale polymorphisms lead to conflicts and breaks in the assembly while conserved regions cross between distant organisms. We present a scaffolding module that aims to address these challenges. It analyzes long-range graph connectivity to identify conserved regions while relying only on local coverage variations. It also identifies regions of polymorphism and avoids fracturing the assembly. We compare our software to current assemblers using simulated and real metagenomic data.

Microbe Assembly Strategies and Finishing at BCM-HGSC

<u>Christian J. Buhay</u>, Yuan-Qing Wu, Shannon Dugan-Rocha, Sandra L. Lee, Michael E. Holder, Yan Ding, Xiang Qin, Vandita Joshi, Irene Newsham, Joseph Petrosino, Sarah K. Highlander, Donna M. Muzny and Richard A. Gibbs

Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030

BCM-HGSC is one of four major centers involved in the Human Microbiome Project. To date, 105 microbial genomes have been assembled as part of the jumpstart portion of this project. 37 of those genomes are currently in the finishing pipeline with 89% completed at various finishing levels. These finishing levels range from automated and or/manual work that provides an "improved" quality over the initial draft sequence to high quality or "gold standard" finished genomes with no unresolved or unconfirmed regions.

As we have continued to evaluate additional microbial genomes, we have identified some primary issues related to assembly and finishing. In addition to informatics based pipeline improvements, experiments with larger 8kb – 10kb insert libraries have been conducted to improve assembly accuracy and contiguity. One major hurdle in finishing microbial genomes has been dealing with large repeats such as rRNA operons. These larger insert libraries enable us to span these repeats and provide better assembly contiguity.

This new technique has been applied to additional microbes currently in the finishing pipeline and further comparisons of 3kb long tag paired ends, 3kb recombi paired ends, and 8kb recombi paired ends for 13 genomes have been completed. Sequence coverage was normalized within each organism as well as across all 13 genomes. Eleven additional 8kb recombi paired-end assemblies were also assembled and analyzed. Assemblies of these microbes with 8kb libraries showed great improvement. More significantly, scaffolding statistics showed that for many of these organisms, most of the genome was found in one large scaffold. For one microbe, *Aeromicrobium marinum* DSM 15272, there was an 8-fold increase in the scaffold N50 from 384 kb in the 3kb long tag paired-end and recombi paired-end assemblies to 3 Mb in the 8kb recombi paired-end assembly. Success with this new technique means most of the genome will be ordered and oriented upon initial assembly, generating higher quality draft assemblies that would require significantly less finishing effort. We are continuing to evaluate and optimize this strategy with additional microbes and plan to implement it for all reference strain sequence assemblies and finishing.

Y Chromosome Finishing Project

<u>Shannon Dugan-Rocha¹</u>, <u>Yan Ding¹</u>, Christian J. Buhay¹, Ziad Khan¹, Michael E. Holder¹, Qiaoyan Wang¹, Wen Liu¹, Jennifer Hughes², Helen Skaletsky², Donna Villasana¹, Lynne Nazereth¹, David Page², Donna M. Muzny¹ and Richard A. Gibbs¹

In an effort to adequately address the Y chromosome and better understand its evolution and biology, the BCM-HGSC is currently sequencing several mammalian Y chromosomes in collaboration with the Page laboratory at the Whitehead Institute. This project specifically targets mammalian Y chromosomes including Macaca mulatta (rhesus monkey), Bos taurus (bovine), and Rattus norvegicus (rat) for additional sequencing and sequence finishing. The Y chromosome is well known for its highly repetitive and largely duplicated regions. To deal with these challenges, traditional Sanger sequencing with a large insert library of 5-8kb has been employed. To date, approximately 415 bovine BACs, 57 macaque BACs, and 74 rat BACs have been prepped, sequenced and assembled in the BCM-HGSC pipeline. Of these, 372 bovine BACs, 56 macaque BACs and 11 rat BACs totaling over 75Mb of unique sequence have been completed at the highest quality or "gold standard".

At an average insert size of 175kb, these difficult BACs with 8-10X Sanger coverage usually assemble into fewer than 10 contigs. Many of these contigs may be misassembled over large duplicated regions and must be manually sorted using read pair information. Results have shown that the difficulty of sorting and binning many of these larger duplications is highly dependent on the number of copies present in the BAC. After tedious tearing and re-joining of contigs, closure of all remaining gaps and low quality regions is largely dependent on direct sequencing of BAC DNA amplified with the GE TempliPhi Sequence Resolver Kit. Although results may vary depending on the size and complexity of the target sequence, results have shown success rates of 95%-98% with an average Phred20 of 550bp. Transposon bombing or specialized shotgun libraries have also been applied for difficult regions such as larger tandem repeats or GC rich regions. Finally, each completed BAC is validated by at least two restriction digestions to confirm assembly contiguity and accuracy. Additional direct BAC sequencing data and results of these strategies will be presented.

¹Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030, ²Howard Hughes Medical Institute, Whitehead Institute, Massachusetts Institute of Technology, Cambridge, MA 02142

FF0109 #1

Sequencing and analysis of virus-CRISPR interactions in a natural system

Christine Sun, Brian Thomas, Jill Banfield

University of California, Berkeley, CA 94720

The recent identification of CRISPRs (clustered regularly interspaced short palindromic repeats) as rapidly evolving, adaptive microbial immune systems provides a new route for studying virus-host interaction dynamics. CRISPRs contain short sequences (spacers) originating from viral or other foreign DNA. After incorporation into the CRISPR region, spacers serve as a template for the targeted destruction of the source DNA. Since spacers are incorporated in a unidirectional manner, CRISPRs serve as historical timelines of virus exposure. The relatively low diversity microbial community in acid mine drainage (AMD) from Iron Mountain, CA, allows the opportunity to develop methods to comprehensively examine the dynamics between virus populations and the CRISPR loci of targeted natural bacterial and archaeal populations.

CRISPRs assembled from AMD community genomic datasets are extremely dynamic, rapidly diversifying regions that display evidence of extensive loss and gain of spacers. Consequently, traditional assembly methods are inadequate for locus reconstruction, necessitating strategies that rely upon analysis of single reads. For studies that use 454 pyrosequencing, high error rates in single reads cause additional challenges. We have developed a pipeline to process reads from large metagenomic and targeted datasets to extract spacer sequences. The spacer sequences, as well as associated metadata (library/sample, read of origin, position on read, unique spacer identifier, etc.) are used to populate a database that enables us to track CRISPR locus dynamics over space and time. The approach has been used to follow virus-host interactions in AMD biofilms and document both static and diversifying loci from different lineages.

Poster Session Notes

Poster Session Notes

FF0109 #2

rBFH artifacts can affect up to 10% of reads in 454 genomic datasets

Vincent J. Denef, Christine L. Sun, Jill F. Banfield

University of California, Berkeley, CA 94720.

We have observed a novel artifact in genomic datasets generated using Roche/454 FLX Titanium technology. The phenomenon resembles the well-known artifact of stacked replicate reads ("bubble reads" or "ghost reads"), which all start at the same location and are presumably created during the emulsion PCR step. Reverse bubble-forming hotspots (rBFH) consist of stacks of up to hundreds of reads that: 1) all appear on the same strand, 2) all end in the same location, and 3) reproducibly form across independent libraries. rBFHs generally occur where other reads (on both strands) read through the region. Yet, in some cases the extensive stacking can lead to dead ends in contigs, resulting in irresolvable gaps in assemblies. In the datasets where we observed this phenomenon, 7-10% of the reads are rBFH artifacts.

We observed an organism independent conserved motif that follows (and sometimes includes) the termination of the rBFH reads. This motif is complementary to the 3' end of the PCR primer within the B adapter. Inspection of the untrimmed rBFH reads revealed that their 3' end contains the B PCR primer, but not the sequencing primer nor the key. This suggests that library fragments containing an internal sequence that is partially complementary to the B PCR primer (specifically to 5'-...GTGCCTTG-3') allow for mispriming during the emulsion PCR. This results in covalently attached fragments that only contain the B PCR primer followed by the truncated library fragment. Preliminary data also indicate that fragments that share more identity with the adapter result in more stacked reads than those that are less similar. While the reason for both mispriming and the enrichment of these sequences over any other fragments in the library is unclear, we hypothesize that kinetic advantages to short fragments during PCR might be at the basis.

DNA Sequencing Pipeline for Contig Auto Assembly and Automated Request Completion

<u>Li, Tony</u>; Hicks, Lyndon; Koffman, Don; Aitzahra, Mostafa; Kieleczawa, Jan; Sookdeo, Hemchand.

Global BioTherapeutics Research, Pfizer

Although many core facilities have embraced next generation DNA sequencing technology since its introduction in 2005, the traditional Sanger method continues to thrive and in many cases expand its usages. Since this technology is mature and robust, the issue is not how to generate good quality chromatograms, but how to efficiently process and analyze the data. For years we have recognized that the true bottleneck in the overall Sanger sequencing pipeline is data finishing and analysis. This is particularly important in the pharmaceutical industry, where quality and error-free data is of paramount importance and where the preferred procedure is still the manual editing of overlapping reads. Frequently, however, the strict adherence to manual editing may not be necessary and more automated protocols can be implemented.

In our lab we have built a custom DNA sequencing LIMS that is exclusively tailored towards finishing work and recently we have added the contig auto assembly (AutoSeq) and auto completion modules (AutoComplete) to our pipeline. The AutoSeq module automatically assembles all reads belonging to any requested DNA, and compares that contig to the reference sequence (if provided). If needed, the output can be reviewed in the Sequencher application and then the finished result is transferred to our LIMS sequencing database and finally posted to the central storage for all sequencing results (WyseCat).

Since the AutoSeq module was implemented few months ago, over 95% of our sequencing reads are processed through this pipeline, resulting in an over 65% time-saving compared to our standard manual finishing effort. The implementation of this workflow enabled us to more than double our sequencing capacity. Currently, we continue to refine AutoSeq and AutoComplete modules thereby providing the ability to process even larger number of sequencing requests with no or minimal human intervention.

Project Management and Targeted Finishing in a High-Throughput Finishing Pipeline

K. Davenport*, L. Meincke, L. Goodwin, O. Chertkov, C. Han, and C. Detter.

Los Alamos National Laboratory, Los Alamos, NM.

New developments in high-throughput finishing include automated start-up, automated closeout, targeted finishing, and project management. With the increasing number of projects coming into our finishing pipeline we are actively developing new ways to streamline the finishing process. We're working to decrease the amount of time a project spends in manual finishing by improving and automating the first phase of genome finishing as well as implementing a formal project management system. These improvements will allow for a faster genome finishing process while reducing costs and improving overall tracking and management of projects at JGI-LANL. Additionally, at a collaborator's request, we can quickly finish targeted regions of interest within a genome while still maintaining the capability to finish the entire genome. We have developed these new targeted finishing strategies in an effort to aid specific and important research interests of the Bioenergy Centers. Several projects have undergone this new approach which has produced the targeted data of interest in a much shorter turnaround time to aid researchers for various bioenergy studies.

Plumbing the hidden depths of the Wellcome Trust Sanger Institute's Illumina Production Pipeline

Ruth Gilderthorp

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA, UK

The Zebrafish sequencing project at the Wellcome Trust Sanger Institute (WTSI) followed a traditional capillary based shotgun sequencing method until July 2009. The progression of second generation sequencing technologies has provided new opportunities for the Zebrafish project. In order to utilise these improvements we are currently sequencing, assembling and finishing pooled Zebrafish bacterial artificial chromosomes (BACs) using combined Illumina GAII data and Whole Genome Shotgun (WGS) capillary reads. Results so far have provided evidence that utilising this sequencing platform can deliver coverage and contig numbers comparable to the capillary method. In order to establish this method a new Illumina Production Pipeline was created. The Pipeline was designed to generate data useful to the Zebrafish Community in addition to providing a starting point for manual finishing. (See Lauren Robertson, Finishing Zebrafish Clones sequenced on Illumina GAII's; old problems on new technology, this meeting)

Individually prepped BAC clone DNA was pooled in units of 12 and was then used to create noPCR libraries, with an insert size of 500 base pairs for the Illumina platform. Utilising noPCR libraries ensures good representation of TA di-nucleotide repeats that are prevalent in Zebrafish by reducing the likelihood of an uneven read distribution (Kozarewa *et al.*, 2009).

The Illumina data produced is combined with WGS capillary reads to generate an assembly with 200 fold coverage. These assemblies are created by Phusion Fuzzy Path (Sudbery *et al*, 2009) and SSAHA (Sequence Search and Alignment by Hashing Algorithm) (Ning *et al*, 2001). Aligning data from WGS capillary shotgun data ensures minimal contig numbers and provides scaffolding information to bridge sequence gaps in the BAC's.

The WTSI production pipeline will be presented. Adaptations we have made from the traditional process together with improvements for the future will be presented.

Finishing Zebrafish Clones sequenced on Illumina GAII's; old problems on new technology.

Lauren Robertson

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA UK

As a consequence of advancements in sequencing technology, the data generation strategy for the Zebrafish project at the Wellcome Trust Sanger Institute (WTSI) changed (See Ruth Gilderthorp, Plumbing the hidden depths of the Wellcome Trust Sanger Institute's Illumina Production Pipeline, this meeting). Incorporated with this change, was the development of an Illumina Production Pipeline which produces draft sequence clones suitable for manual finishing. By combining data from Illumina GAII and Whole Genome Shotgun (WGS) capillary reads this method can produce assemblies comparable to that of traditional capillary sequencing allowing finishing to a Phase 3 accession of the HTGS division of EMBL.

BAC (bacterial artificial chromosomes) clone DNA, which is pooled in groups of 12, is used to create noPCR libraries for the Illumina platform. Data generated is assembled using Phusion Fuzzy Path (Sudbery et al., 2009) which results on average, 120 contigs with a coverage of greater than 95%. Due to the assembly algorithm utilized gaps are still present in the data, hence prior to employing additional chemistry, IMAGE (Iterative Mapping and Assembly for Gap Elimination) (Tsai) is run. By sequence searching contig ends, IMAGE can find additional data to extend contigs and close gaps, closing approximately 55% of gaps. Following this, an initial round of automated PCR selection is generated using ABACAS (Algorithm Based Automatic Contiguation of Assembled Sequences) (Assefa et al., 2009). ABACAS makes use of a reference sequence and uses this to align, order and orientate data. Furthermore, it can automatically generate primer oligos for gap closure using Primer3 (Rozen and Skaletsky 2000). From here finishing is carried out adhering to many of the standards already in place at the WTSI.

Visualization and manipulation of these clones is currently carried out using Gap4 (Staden 1995) but with a view to eventually switching to Gap5 (Bonfield), currently under development. When completed, Gap5 will permit users to view individual Illumina reads and confidence values associated with these allowing a better understanding into how the Illumina and capillary data contribute to the final consensus.

Here we present the finishing strategy that has been formulated examining 72 BACs that were combined in 6 pools and delivered finished clones.

Echinococcus Multilocularis Sequencing Project

Anne Babbage, on behalf of the Helminth Finishing groups.

Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK

The Wellcome Trust Sanger Institute (WTSI) is currently working on one of the first cestodes to be sequenced. Echinococcus multilocularis is a fox tapeworm, the larval stage of which causes alveolar echinococcosis in humans and in most cases is fatal. Increased disease prevalence in humans is thought to be linked to the migration of wild foxes to urban areas. This cestode is relatively easy to cultivate in the laboratory, with rats being the intermediate host, making it a model subject for a sequencing study. E.multilocularis is closely related to E.granulosus (hydatid disease) and Taenia solium (pig tapeworm) which are both important pathogens, therefore it will serve as a very useful reference for future cestode sequencing projects. The genome size is estimated to be 120MB across 9 chromosomes (2n=18). It is the first large scale sequencing project at the Sanger Institute to take advantage of three sequencing technologies: illumina (150X), Roche Life Sciences 454 (9X) and capilliary (6X). The assembly was made using shredded consensus from 454 reads, capillary shotgun, fosmid end sequences and capillary shotgun and pre-finishing reads. In house program IMAGE (Iteratively Mapping and Assembly for Gap Elimination) was used to close gaps using illumina reads (818 closed with 700 bridged gaps remaining). Another in house program ICORN (Iterative Correction Of Reference Nucleotides) was then used to highlight possible sequence errors, regions of extreme illumina coverage and possible heterozygous bases to improve the assembly and aid manual finishing (5064 regions tagged).

Here we present an overview of the genome project, including current progress and statistics. We will also illustrate future strategies such as use of transcriptomics data.

Evaluation of Scaffold Gap Closure Methods Using 454 Sequencing

<u>Chelsea Dunbar</u>, Anna Montmayeur, Amr Abouelleil, Harindra Arachchi, Gary Gearin, Annie Lui, J. Pendexter Macdonald, Margaret Priest, Niall Lennon, Jennifer Baldwin, Bruce Birren and Michael G. FitzGerald

Broad Institute of MIT and Harvard, Cambridge, MA

The study of emergent pathogens has changed substantially with the advent of next generation sequencing platforms. Streamlined sample prep, instrument sequencing capacity and effective assembly and read mapping software facilitate our ability to study pathogen outbreaks using whole genome sequence. We are now able to study topics such as bacterial transmission and the rise of drug resistance in exquisite detail.

Reference quality, finished sequence is an important component of these efforts. Complete genomes provide a comprehensive template for rapid read mapping along with the organism's entire gene complement. While the next-gen platforms have dramatically accelerated our ability to produce draft sequence, similar gain in our genome finishing capacity is yet to be realized. One approach to accelerate the elucidation of directed sequence involves taking advantage of the next generation sequencing platform's high sequencing capacity.

The lack of physical clones complicates gap closure, dictating that bacterial genome finishing largely remains the domain of PCR and Sanger sequencing. Closure of scaffold gaps, those not spanned even by a next-gen jumping read, is a critical part of the finishing process. We will detail Broad efforts to improve the efficiency of large and scaffold gap closure. In particular, we are developing a series of approaches using *Enterococcus casseliflavus* 899205, *Mycobacterium tuberculosis* H37Rv, and *Staphylococcus aureus* 55/2053 as test templates and all for eventual sequencing on the Roche-454 system. Standard PCR, combinatorial PCR and bubble-adapter PCR are used to generate sequencing template. Multiple sequencing approaches are under evaluation. One variant employs shear library preparation from the larger molecules. We have also employed end sequencing, including from the adapter end of bubble-adapter amplicons. Molecular bar codes are utilized to maximize PTP efficiency. We will present results on 454 sequenced targets versus Sanger controls.

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900006C.

Multiplex Template Preparation for Finishing With Next Generation Sequencing

<u>Anna Montmayeu</u>r, Gary Gearin, Amr Abouelleil, Annie Lui, Niall Lennon, Rachel Erlich, Harindra M. Arachchi, Chelsea Dunbar, Jennifer Baldwin, Bruce W. Birren and Michael G. FitzGerald

Broad Institute of MIT and Harvard, Cambridge, MA

Finished genome sequence is an important component of our bacterial pathogen research program. Complete sequence facilitates identification of resistance mechanisms, epidemiological work and provides a complete reference for mapping related strains.

Finishing by means of next generation sequencing technologies is only cost effective when a very large number of targets are processed simultaneously. In the case of the Roche-454 GS FLX Titanium sequencer this calls for the labor-intensive steps of large-scale PCR amplicon generation followed by sample barcoding and 454-adapter addition, so that samples can be pooled together prior to sequencing.

In an effort to simplify sample preparation we turned to molecular inversion probes (MIPs), also known as "padlock probes". MIPs were originally used for high-throughput, multiplex SNP detection. Recently they have also been used for the multiplexed capture and sequencing of a large number of small exons.

We are conducting a series of pilot experiments using pathogen genomes currently involved in our active finishing process. Targets are designed around gap regions of the pathogens, *Neisseria gonorrhoeae* MS11, *Enterococcus casseliflavus* 899205 and *Staphyloccus aureus* 55/2053. All assemblies are based on 454-FLX data and are assembled using Newbler. We will present an evaluation of MIP results for 64 gaps. All have reference data available from our production (Sanger) finishing process. Pilot targets range from 0.2kb to 2.5kb. Preliminary results are quite positive and demonstrate our ability to capture and derive sequence from targets up to 1.5kb. This project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases National Institutes of Health, Department of

Health and Human Services, under Contract No. HHSN272200900006C.

Origins of broadly neutralizing anti-HIV-1 antibodies as elucidated by deep sequencing of expressed immunoglobulin heavy chain repertoires

Henry KA1, Murira A1, Lepik C1, Breden F2, Bradbury ARM3, Scott JK1

Departments of 1Molecular Biology and Biochemistry and 2Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6 3Bioscience division, Los Alamos National Laboratory, Los Alamos, NM 87545

Antibodies (Abs) capable of neutralizing a broad spectrum of heterologous HIV-1 isolates are rarely produced in natural infection, yet their elicitation will likely be a requirement for a successful vaccine. Several of the monoclonal broadlyneutralizing (bNt) anti-HIV-1 Abs identified to date have very long third complementarity-determining regions of the heavy-chain (CDRH3s) and high levels of somatic mutation (SM). Furthermore, it has been proposed that they are polyreactive autoAbs and thus unlikely to be elicited by conventional vaccination approaches. We have assembled a sequence database of 590 human MAbs of known specificity isolated from patients with chronic viral infections, acute viral or bacterial infections, and systemic autoimmune diseases. Comparisons of CDRH3 length, SM rate and VH gene usage reveal that bNt anti-HIV-1 MAbs are indistinguishable from anti-HIV-1 MAbs as a whole, and that long CDRH3s, high SM and distal VH gene usage may be general features of T-cell driven Ab responses arising in chronic viral infection. In contrast, Abs from acute viral infection have short CDRH3s and low levels of SM, while Abs arising in autoimmunity have intermediate CDRH3 lengths and SM rates. We now plan to analyze sera and peripheral blood lymphocytes from HIV+ donors whose sera are bNt, HIV+ donors whose sera are not bNt, patients with systemic lupus erythematosus, and healthy donors. Peripheral B cells will be sorted into naïve, memory and plasmablast subsets, and the full repertoire of expressed VH genes from each cellular pool will be amplified by RT-PCR, sequenced to full coverage, and analyzed for VH gene usage, CDRH3 length and SMs. Sera will also be analyzed on protein microarrays bearing autoantigens, HIV-1 antigens, and markers of polyreactivity. We expect to describe the specific molecular features associated with broad neutralization, which should better guide HIV vaccine design.

Gap Closing/Finishing by Targeted Genomic Region Enrichment and Sequencing

Kanwar Singh, Jeff L. Froula, Hope Tice, Len A. Pennacchio, and Feng Chen

Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California, United States.

* This work is done in collaboration with RainDance Technologies*

Gap Closing/Finishing of the draft assemblies of genomes is a labor and cost intensive process whereby several rounds of repetitious amplification and sequencing are carried out. Here we demonstrate a high throughput procedure where custom primers flanking gaps in draft assembled genomes are designed and primer libraries containing up to 4,000 unique primer pairs in independent droplets are prepared. Then by using RainStorm microdroplet-based technology from RainDance, droplets of custom primer library and fragmented genomic DNA template are merged to form millions of picoliter scale droplets. Each droplet is the functional equivalent of an individual test tube, which are pooled and amplified. The resultant enriched fragments are concatenated and processed through Illumina sequencing. The Sequencing data is assembled and resulting contigs can be added to draft assembly to close gaps. We will present overall experimental strategy, primer design algorithm and initial results.

Functional Data Analysis Approach to Association Studies of Genetic Data from Next-generation Sequencing

Li Luo, Eric Boerwinkle and Momiao Xiong

Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston

Genome-wide association studies (GWAS) have become the primary approach for identifying genes with common variants influencing complex diseases. Despite considerable progress that GWAS have made in identifying hundreds of putative disease loci, these newly discovered loci account for only a small proportion of disease heritability. A potential source of the missing heritability is the contribution of rare variants. Millions of rare variants will be discovered by next-generation sequencing technologies. These technologies have three defining features: large number of rare variants, a high proportion of sequence errors, as well as a large proportion of missing data. These features raise great challenges for testing the association of rare variants with complex disease. To meet the great challenges raised by next-generation sequencing, instead of modeling the genome as a few separated individual loci, we use a genome continuum model and functional principal components as a general principle for developing novel and powerful association analysis methods designed for resequencing data. To evaluate the performance of functional principal component analysis (FCPA)-based statistics, we use simulations to calculate the type I error rates and the power of five alternative statistics: FPCA-based statistic, the generalized T², Collapsing method, CMC method and individual χ^2 test. We also examined the impact of sequence errors on the type I error rates of the five test statistics based on 1000 genome data. Finally, we apply the five statistics to published resequencing dataset from ANGPTL4 in the Dallas Heart Study. We report that FPCA-based statistics have higher power to detect the association of rare variants and stronger ability to filter sequence errors than the other four methods. Our work represents a paradigm shift from the current single marker association analysis to sequence-based association analysis.

Overcoming Some of the Challenges to Single Cell Genomics

<u>Janey Lee¹</u>, Damon Tighe¹, Mei Wang¹, Stephanie Malfatti¹, Mansi Chovatia¹, Shweta Deshpande¹, Erika Lindquist¹, Feng Chen¹, Ramunas Stepanauskas², Jan-Fang Cheng¹ and Tanja Woyke¹

¹DOE Joint Genome Institute, Walnut Creek, California and ²Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine

Single cell genomics, the amplification and sequencing of genomes from single cells, can provide a glimpse into the genetic make-up and thus life style of the vast majority of uncultured microbial cells, making it an immensely powerful and increasingly popular tool. This is accomplished by use of multiple displacement amplification (MDA), which can generate billions of copies of a single bacterial genome producing microgramrange DNA required for shotgun sequencing. Here, we would like to address several challenges inherent in such a sensitive method and propose solutions for the improved recovery of single cell genomes. While DNA-free reagents for the amplification of a single cell genome are a prerequisite for successful single cell sequencing and analysis, DNA contamination has been detected in various reagents, which poses a considerable challenge. Our study demonstrates the effect of UV radiation in efficient elimination of exogenous contaminant DNA found in MDA reagents, while maintaining Phi29 activity. Second, MDA is subject to amplification bias, resulting in uneven and sometimes insufficient sequence coverage across the genome. In a post-amplification method, we employed a normalization step within library construction in both 454 Titanium and Illumina platforms. Populations of highly abundant sequences were specifically targeted and degraded from the library via duplex-specific nuclease, resulting in decreased variability in genome coverage. While additional challenges in single cell genomics remain to be resolved, the two proposed methodologies are relatively guick and simple and we believe that their application will be of high value for future single cell sequencing projects.

Use of Optical Mapping to Aid in the Assembly and Finishing of Human Microbiome Genome Projects

T. K. Wagner, E. A. Meudt, E.B. Zentz. OpGen Inc., Gaithersburg, MD 20878.

One of the goals of the Human Microbiome Project (HMP) is to sequence the whole genomes of over 1,000 human commensals. However, currently less than 2% (25/1482) are considered as contiguous and finished. Optical Mapping is a technology that can rapidly yield whole-genome ordered restriction maps that can be used as a whole-genome scaffold to position, order, orient, and validate the assembly of DNA sequencing contigs. The hypothesis of this study was that Optical Mapping could be used to place the majority of contigs 40 kb and larger of the selected HMP reference genomes.

Whole-genome Optical Maps were generated from 7 HMP reference microbes ordered from ATCC, which included *Chryseobacterium gleum* (ATCC 35910), *Citrobacter youngae* (ATCC 29220), *Enterobacter cancerogenus* (ATCC 35316), *Neisseria cinerea* (ATCC 14685), *Neisseria sicca* (ATCC 29256), *Streptococcus infantarius infantarius* (ATCC BAA-102), and *Weissella paramesenteroides* (ATCC 33313). Contigs of wholegenome DNA sequence for each bacteria were obtained from GenBank, imported into MapSolver™ software to generate an *in silico* restriction map of each contig, and compared to the Optical Maps.

Over 85% of the contigs 40 kb and larger were positioned and ordered against the Optical Maps. The placement of contigs resulted in an average of over 75% coverage of the unfinished genomes. The gaps between contigs were located and the sizes determined. Potential misassemblies were identified in the DNA sequence contigs.

These data support the ability of Optical Mapping to position, order, orient, and validate the assembly of DNA sequencing contigs for whole-genome sequencing projects. The use of Optical Mapping could not only help assemble the 98% of the HMP reference genomes that are still considered unfinished but also decrease the time to do so by locating gaps and identifying misassemblies in the DNA sequence contigs.

Clone-based Sequencing Using Next Generation Technology

Tina Graves

The Genome Center/Washington University School of Medicine

With increased ease and lower cost of sequencing whole genomes with new sequencing technologies, there is a decreased need for clone-based projects. In many cases it is possible to produce a final high quality product without clone-based efforts, however there are some projects where clone-based methods are still preferred. One such project is currently underway through the Genome Reference Consortium (GRC). The goal of the GRC is to correct the small number of regions in the human and mouse reference sequences that are currently misrepresented, due to errors and gaps as well as provide alternative assemblies of structurally variant loci. In order to adhere to the standards previously set for the current reference sequence, clone-based sequence is being used where possible. These regions of the genomes, particularly the gaps and structural variation, tend to be associated with repetitive sequences which make them better suited for the traditionally longer paired end sequences that result from capillary Another project requiring clone-based effort is the NHGRI Structural sequence. Variation Proiect (http://www.ncbi.nlm.nih.gov/projects/genome/StructuralVariation/NHGRIStructuralVaria tion.shtml). For this project, the goal is to sequence new euchromatic sequences that are not represented in the reference due to structural variation. Again, these new sequences typically tend to be associated with duplication and repeat rich genomic intervals, thus the whole genome approach is less effective than clone-based methods.

Until recently, clone-based sequencing projects have utilized capillary sequencing platforms, with libraries being created on a clone-by-clone basis. This method is costly as well as labor intensive. Because of this, and the need to continue working on a small number of clone-based projects, we have developed methods to sequence a pool of clones using the Roche 454 sequencer. In comparing costs of the two methods, a 4fold cost reduction is realized with the next generation sequencing method. Although there are significant cost savings using this method, there are some regions of the genome that do not lend themselves to sequencing with only next generation For these cases, we are investigating combined capillary and next generation sequence data, utilizing the strengths of each platform. In order to increase the efficiency of generating the capillary data, the library construction event will be done on a pool of clones. With this strategy a 2-fold decrease in cost will be realized when compared to the strategy of capillary only data on a clone-by-clone basis. Utilization of next generation technology where possible, enables continued clone-based work in a much more efficient manner, while still providing the benefits of localized clone-based assembly.

Improving Structural Variation Detection with De Novo Assembly of NGS reads

Joel Martin, Kurt LaButti, Alla Lapidus, Igor Grigoriev, Len Pennacchio, <u>Wendy</u> Schackwitz

U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598 USA

We have analyzed the efficacy of local de novo assembly, using a variety of assemblers, for improving and confirming structural variation detection at varying depths in a set of microbes, fungi, and plants. Additionally we compare local de novo with whole genome de novo across a subset of the organisms.

Short-read Assembly of Bacterial and Archaeal Genomes

Alicia Clum, Brian Foster, Kurt LaButti, Alexander Sczyrba, Alla Lapidus and Tanja Woyke

DOE Joint Genome Institute, Walnut Creek, California 94598, USA

Since the emerging of second generation sequencing technologies, the evaluation of different sequencing approaches and their assembly strategies for different types of genomes has become an important undertaken. Next generation sequencing technologies dramatically increase sequence throughput while decreasing the cost, making them an attractive tool for whole genome shotgun sequencing.

To compare different approaches for de-novo genome assembly, appropriate tools and a solid understanding of both quantity and quality of the underlying sequence data are crucial. Here, we performed an in-depth analysis of short-read Illumina sequence assembly strategies for bacterial and archaeal genomes. Different types of Illumina libraries as well as different trim parameters and assemblers were evaluated.

Assemblies were generated using Velvet (publicly available assembler) and MERaculous (JGIs in house development). Results of the comparative analysis of the tools and sequencing platforms will be presented.

The goal of this analysis is to develop a cost-effective approach for the increased throughput of the generation of high quality microbial genomes.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Controlling The Variation in Microbial SNP Detection

Sean M. Sykes, Sarah Young, Aaron Berlin, Carsten Russ and Chad Nusbaum

Broad Institute of MIT and Harvard, Cambridge, MA

The ability to assess microbial sequence diversity has increased at an unprecedented level thanks to the low-cost, high-yield next generation sequencing platforms. It is crucial to be able to accurately and reliably curate these data to uncover the underlying sequence polymorphism that yields increased drug resistance or pathogenicity. The high quality, deep sequence coverage generated by the Illumina sequencer provides data that enables the ability to detect mutations with confidence, however, this method faces challenges both new and old from short read lengths, base composition bias, repeat content, and availability of a complete reference sequence.

We have devised a framework to rationally, fairly, and objectively evaluate and compare alignment and SNP detection algorithms. We began by generating controlled sets of *in silico* mutations along a finished reference and examining how alignment sensitivity is affected by increasing variation. The truth set allows for accurate evaluation of SNP detection through True Positive, False Positive, and False Negative rates. Using these techniques we have outlined a protocol for quickly, accurately, and confidently assessing microbial diversity. Our target datasets use genomes for which finished sequence is available to assess performance and include a range of GC content and genome size. These test sets include *E. coli, Rhodobacter sphaeroides* (high GC) *Plasmodium falciparum* (low GC), *S. pombe*, and ultimately large genomes as well.

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900006C

Telomere Assembly: The Means To An End

David Heiman, Aaron Berlin, Sean Sykes, Sarah Young, and Chad Nusbaum

Broad Institute, 320 Charles St. Cambridge MA

Due to the repetitive nature of the underlying sequence (and perhaps to their location at the ends of chromosomes), telomeres are difficult to capture and sequence by traditional clone—based methods, and a challenge to assemble. New sequencing technology data and our new computational approach offer a solution to this challenge. We have developed a process to discover and assemble the telomeric sequences of fungal genomes using a variety of sequencing technologies. There are two major stages of this process: finding the telomeric repeat elements along with the reads that contain them (along with their mate-pairs), and putting these reads together into an analyzable telomere assembly.

Inputs to the process are a draft assembly of the genome and a set of WGS reads (which are not required to be included in the assembly). Finding the telomere begins with a *de novo* analysis of the reads for tandem repeats with units that fall within the length of potential telomere motif. If any experimental data are available (such as a restriction map), they are used in conjunction with the *de novo* analysis to confirm the telomere motif. Once the likely telomere sequences are identified, an iterative assembly process is used to capture additional putative telomeric reads and extend the telomere assembly. If paired reads are available, this process involves anchoring read mates that do not contain the telomeric repeat to the scaffold ends that abut telomeres in an assisted assembly to begin building out the telomeres, aligning the entire read set to the assembly, and reassembling the enhanced read set.

The process can involve the use of several assemblers, depending on the types of datasets available. Notably, HybridAssemble from the Arachne suite for anchoring and extension using ABI 3730 reads, Newbler for assembling and mapping 454 reads, and ABySS to put together a variety of data types.

Schizosaccharomyces octosporus and S. japonicus are being used as example organisms with which to develop this process. Both have a variety of read types available, including long read pairs and new sequencing technology data, as well as restriction maps that can help corroborate any findings.

High Throughput Laboratory Network--- An Application of High Throughput Sequencing and Analysis

<u>Helen Cui</u>, Tracy Erkkila, Lance Green, Cheryl Gleasner, Jennifer Harris, Chris Detter, Los Alamos National Laboratory

The public suffers daily from infectious disease impacts to human health, agriculture and the environment. Global infectious disease surveillance has been recognized as an important approach to combat infectious diseases, to effectively protect public health and global security. The Los Alamos National Laboratory (LANL) and the University of California at Los Angeles (UCLA) are collaborating to develop a High Throughput Laboratory Network (HTLN) aiming to establish high-speed, high-volume laboratory capabilities, a technology enabling global surveillance for early detection and accurate analysis of pathogenic agents.

The HTLN will automate a workflow that includes sample collection, laboratory testing, and data management and analysis. The architecture of the HTLN is modular in nature and consists of systems for sampling, accessioning, extraction, screening culturing, genotyping, and phenotyping. Overall, the HTLN is designed to process 3,000 viral samples per day in rapid screening mode, and up to 10,000 full virus genomes per year. The sample accessioning system and biobank have been completed. Automated sample screening and pathogen culturing systems are under procurement, and a pathogenic phenotype analysis system is being designed. The genotyping capability of the HTLN currently uses Sanger chemistry. Starting with RNA, multiple primer sets are used to amplify entire genomes in small overlapping segments. All steps in the process are completely automated on the system. With current primer sets, up to 160 samples can be sequenced in an 11.5 hour run. UCLA will operate the first HTLN lab and serve as a model. The anticipated laboratory network will function much like a parallel high performance distributed computer in quickly solving a problem. The system will be housed in Biosafety Level 3-enhanced containment to enable the processing of samples containing highly pathogenic avian influenza viruses and other potential select agents.

A General Method for Assembling Genomes from Illumina data

<u>Dariusz Przybylski</u>, Iain A. MacCallum, Sante Gnerre, Joshua Burton, Filipe Ribeiro, Bruce Walker, Ted Sharpe, Giles Hall, Genome Sequencing Platform, Carsten Russ, Chad Nusbaum, David B. Jaffe

Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA

Over the past decade, DNA sequencing costs have dropped about 10,000-fold. The lowest-cost reads from current technologies are short, have a high error rate, and land unevenly on the genome. They are ideally suited to 'resequencing' applications in which a preexisting reference sequence is available; but for *de novo* genome assembly, they are challenging to work with, and results have generally been inferior to those obtained from the old (Sanger method) technology.

Here we demonstrate a practical and general laboratory/computational method for generating high-quality *de novo* assemblies of genomes at the lowest possible cost. Our method starts with 100-base Illumina paired reads from two libraries: one from fragments of size 180 bp (slightly less than twice the read length), and one from fragments of size 3000 bp, via a 'jumping' construction. These two libraries use off-the-shelf methods and provide power that could not be obtained from a single library. We also demonstrate experimental methods for jumping longer fragments to yield a third library, providing even greater potential for long-range connectivity.

We assembled these data using our new version of the ALLPATHS algorithm. This algorithm has been scaled up to work on large genomes and made robust to idiosyncrasies in library construction, variation in coverage, and run-to-run variability in sequence quality, all of which have been critical problems for genome assembly.

We tested our method using a suite of 16 genomes, 9 for which a reference sequence was available and 7 from completely new samples. These genomes range in size from 2 Mb to 2.6 Gb, and in GC content from 19% to 71%. Using the preexisting reference sequences, we assess the completeness, continuity, and accuracy of these assemblies, finding that for the smaller genomes their quality exceeds the general quality level of draft assemblies that had been achieved using Sanger sequencing. For the largest (bushbaby), we obtained contigs and scaffolds having N50 sizes 18 kb and 3.9 Mb, respectively. We compared to a preexisting 3.5x Sanger assembly, and by so doing estimated the *combined* misjoin rate in the two assemblies: one per 10 Mb.

Bacterial High-throughput Transcriptomics, Present and Future

<u>Loren Hauser</u>¹, Daniel Quest¹, Andrey Gorin¹, Shihui Yang¹, Steve Brown¹, Sara Blumer-Schuette², Mike Adams³, Bob Cottingham¹, Bob Kelly²

1) Oak Ridge National Laboratory; 2) North Carolina State University; 3) University of Georgia, Athens

High-throughput Transcriptomics (RNAseq) has the potential to dramatically improve the knowledge acquired from systems biology experiments. Generating sufficient RNAseq data is already feasible, and will become considerably faster, increasingly cost effective and just as accurate, as third generation sequencing technologies come online. RNAseg has substantially better dynamic range (10⁶ vs10³) than traditional gene expression arrays, and exquisite sensitivity (1 mRNA molecule/10³ cells). provides direct sequence data for determining operon structure and finding new If collected in sufficient quantity, and combined with additional regulatory RNAs. systems biology data and a comprehensive bioinformatic toolkit, it can be the basis for generating Genetic Regulatory Network predictions. In order to implement this vision, we are developing an automated analysis pipeline for handling RNAseq data that will eventually be able to add some of the information acquired (operons, putative regulatory RNAs, corrected annotations) into our automated genome annotation pipeline and GenBank, as well as initiate the development of a Genetic Regulatory Network Prediction system.

Carrier Screening of Recessive Genetic Disorders by Next Generation Sequencing

<u>D.L. Dinwiddie</u>, C.J. Bell, R.W. Kim, N.A. Miller, B.J. Rice, J.A. Crow, E.E. Ganusova, S.L. Hateley, S.F. Kingsmore

National Center for Genome Resources, Santa Fe, New Mexico

Background and Objective: Human recessive genetic diseases are individually rare, but collectively cause significant morbidity and mortality (20-30% of infant deaths and 11% of pediatric hospital admissions). NCGR is developing a carrier screening test for 448 recessive disorders caused by tens of thousands of mutations.

Methods: Target enrichment of genes, multiplexed deep sequencing, and automated bioinformatic analysis are used to determine carrier status. Two enrichment techniques (RainDance RainStorm and Agilent SureSelect) were evaluated for a series of analytic metrics. Illumina[®] GA IIx and Life Technologies SOliD™ 3 sequencing were evaluated for heterozygosity detection.

Results: On average, 30% & 23% of sequenced nucleotides were on target for RainDance & Agilent, corresponding to fold-enrichment of 462 and 360, respectively. The Alpheus[®] variant detection pipeline correctly identified SNPs and in/dels, zygosity, and carrier status. Illumina and SOLiD™ sequencing of SureSelect™ enriched samples had high concordance in variant and zygosity calls.

Discussion and Conclusions: Carrier status for rare recessive disorders was accurately identified by target enrichment of approximately 2Mb of genomic loci with either Agilent SureSelect or RainDance RainStorm™ technologies followed by Illumina GA IIx or SOLiD™ 3 next-generation sequencing using the Alpheus[®] Sequence Variant Detection pipeline.

This work was supported by the Beyond Batten Disease Foundation.

Next-Generation Transcriptome Assembly of Medicinal Plants – Disentangling Plant Natural Product Pathways

<u>Shannon L. Hateley</u>¹, Arvind K. Bharti¹, John A. Crow¹, Joann Mudge¹, Andrew D. Farmer¹, Robin Kramer¹, Meinhart Zenk², Toni Kutchan², Gregory D. May¹

Medicinal plants have been prized for their healing and preventative properties for centuries. Still today, 25% of pharmaceuticals are derived directly from, or are based upon, natural products. Yet knowledge of the biochemical pathways to these medicinals remains rudimentary. Using Illumina RNA-Seg technology, we are sequencing the transcriptomes of 30 medically relevant plant species to study their medicinal compounds and elaborate their biochemical pathways. Many of these species have little or no available transcriptomic or genomic sequence. Therefore, we are creating de novo assemblies of our Illumina sequence data and any other available transcriptomic data. Sequencing of multiple tissues per species has allowed differential expression and pathway convergence analyses using our Alpheus Bioinformatics Platform and JMP Genomics (SAS). To date, 36 tissues from 10 species have been sequenced and differential expression has been analyzed in two species. Sequence motif discovery using MEME/MAST and HMMER to discover novel Cytochrome P450s has been applied to eight species. Our collaborating biochemists will apply this knowledge to the understanding of plant-derived medicinal compounds and biochemical pathways. This will lead to the development of alternative sources of medically active compounds and the creation of novel drugs.

¹National Center for Genome Resources, Santa Fe, NM

²Donald Danforth Plant Science Center, St. Louis, MO

Predictive profiles of sepsis death and outcomes identified by an omics-based approach

<u>Jennifer C van Velkinburgh, PhD*</u>, Raymond J Langley, PhD, Stephen F Kingsmore, M.B., C.h.B., B.A.O

National Center for Genome Resources, Santa Fe, New Mexico

The Community-Acquired Pneumonia and Sepsis Outcomes Diagnostics (CAPSOD) project is a multidisciplinary collaboration involving investigators at ten organizations. Since 2005, we have been prospectively enrolling a diverse geographical, socioeconomic and ethnic population of patients (n=1000) with suspected sepsis. Sepsis survivor and death profiles have been found to be minimally influenced by progression to severe sepsis or septic shock, etiologic agent or therapy, suggesting that dichotomous host responses to pathogenic infections determine patient outcome. Plasma samples (n>200; ED presentation, to and 24 hours later, t24) were analyzed for specific changes in the metabolite and protein profiles to identify a distinct diagnostic signature that will predict outcomes in sepsis. Predictive biomarker classifiers were identified (t₀) and replicated (t₂₄) that distinguished non-infected, SIRS-positive patients, sepsis survivors and sepsis deaths and that may allow prompt differentiation and individualized treatment of sepsis. Furthermore, we carried out Next-Gen sequencing with Illumina technology on whole blood samples (n=135; t₀) from well-characterized individuals with community-acquired sepsis or non-infected, SIRS-positive controls. A distinct pan transcriptional response was identified in sepsis survivors, sepsis death, and controls, allowing for a comprehensive genetic profile of sepsis progression and outcomes to be constructed. When the comprehensive datasets of differentially expressed genes were analyzed to determine related biological pathways and functions, striking correlations with neutrophilic responses were identified, including perturbations in the inflammatory response, thrombic-fibrinolytic response, oxidative burst and immunoproteosome pathways. In summary, sepsis patients who would die presented significantly decreased transcription of 1326 genes in blood, many elevated metabolic intermediates in plasma and lacked fever, suggesting aborted catabolism of energetic substrates.

*Corresponding author: National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, NM 87505; tel: 001-505-995-4423; fax: 001-995-4461; e-mail: jcv@ncgr.org

Analyzing Cancer at Single Cell Resolution with Droplet Technology

Lin Pham, M.D. Commercial Application Scientist

Co-Authors: Martina Medkova, Eric Brouzes, Elodie Dahan, Steve Kotsopoulos, Jason Warner Jonathan Larson, Jeff Olsen, Darren Link, and Michael Samuels

RainDance Technologies, Inc., Lexington, MA

Cancerous cell growth can result from mutation of the underlying genome or from aberrant epigenomic regulation, altering transcript and protein expression levels and causing uncontrolled growth. The resulting tumors are often composed of heterogeneous mixtures of both clonally transformed tumor cells and other cell types. Here we describe extensions of our novel microdroplet technology that enable analysis of a DNA sample's methylome and allow detection of allelic variants in heterogeneous mixtures. We also present unique new capabilities for capturing both phenotypic and genomic information from droplet-encapsulated individual cells, enabling analysis of tumors with single-cell resolution.

RainDance Technologies' microfluidic technology produces uniform picoliter-scale aqueous microdroplets at rates up to 10 million droplets per hour. Each droplet is the functional equivalent of an individual test tube and can contain a single molecule, reaction, or cell. This versatile technology can adapt proven assays to high-speed workflows with a minimum of process-induced bias or errors. Our initial application provides targeted sequence enrichment to prepare samples for next-generation sequencing (illustrated below). The sequencing depth and reduced amplification bias provided by microdroplet technology enables accurate detection of sequence variants within a heterogeneous mix of sample DNA. In addition, we have extended this approach to enable targeted sequencing of a sample's methylome using bisulfite-treated templates, providing base-pair resolution of the methylated cytosines that have been associated with aberrant transcription in cancer.

The self-contained microdroplet environment can also help scientists study biological samples as a collection of individual cells. Single-cell droplet technology provides a cost-effective method to gain sequence information from individual cells that have been sorted by phenotype. We present data illustrating the workflow we are using to perform low-bias single-cell whole-genome or transcriptome amplification within the sorted microdroplets. The amplified single cell genomes can be used as the starting material for targeted sequence enrichment using the RDT1000, providing a automated sample preparation method for single cell analysis using any next-generation sequencing platform. Single-cell droplet technology is an exciting new tool that will allow for a more thorough examination of the variations that influence the predisposition and composition of cancer, and the cellular responses to therapeutic and prevention agents.

Sensitive Detection of Rare p199RY Escape Variants via Ultradeep Sequencing

<u>Peter Hraber</u>1, Evan Cale2, Elena Giorgi1, Will Fischer1, Thomas Leitner1, Cheryl Gleasner3, Lance D. Green3, Cliff S. Han3, Norm Letvin2, Bette Korber1

- 1 Theoretical Biology & Biophysics, Los Alamos National Laboratory, Los Alamos NM 87545.
- 2 Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, Boston, MA 02215
- 3 DOE Joint Genome Institute, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545.

In simian immunodeficiency virus (SIV), p199RY is an immunodominant epitope restricted by Mamu-A*02 alleles. Previous work has reported cytotoxic T lymphocyte (CTL) response to escape mutations in this epitope by 35 days post-infection (dpi), though this is earlier than epitope variants had been detected via conventional We used ultradeep 454 pyrosequencing of nef amplicons spanning p199RY to identify rare escape variants and characterize escape dynamics over time. In 5 Mamu-A*01+/A*02+ rhesus macaques intravenously infected with 60,000 copies per ml of SIVmac251 inoculum we find the 2 most abundant sequences (>95% of sample) 21 dpi differ only by a polymorphic silent substitution outside the epitope. A Poisson model of sequence divergence indicates multiple founder viruses are present or selection is active at 21 dpi. Amino-acid diversity is no greater inside the epitope than outside the epitope at 21 dpi, though epitope sites are more variable at 35 or 84 dpi (one-sided Wilcoxon p<0.05). The transmitted epitope predominates (>99.8% frequency), yet rare escape variants occur in all 5 animals at 21 dpi, which outcompete all others by 84 dpi. Together, our results suggest that the earliest escape mutations occur de novo in the host, rather than transmitted from the inoculum. Therefore, CTL selection for SIV epitope escape variants is extremely rapid, and active earlier than 21 dpi. Ultradeep sequencing provides great sensitivity to detect rare variants and reveals that CTL activity influences virus evolution early in infection, before peak CTL response, and earlier than previously detected via conventional sequencing.

Funding: Laboratory Directed Research & Development (LDRD), Center for HIV/Aids Vaccine Immunology, National Institute of Allergy and Infectious Diseases

Meet and Greet Party

600pm - 830pm, June 2nd

Sponsored by Roche Diagnostics

Enjoy!!!



06/03/2010 - Thursday				
Time	Туре	Abstract #	Title	Speaker
7:30 - 8:30am	Breakfast	x	Santa Fe Breakfast Buffet Frittata with applewood smoked bacon and English muffin, Fruit jam and butter, Freshly brewed regular and decaffeinated coffee	х
8:30 - 8:45	Intro	х	Welcome Back from DOE Joint Genome Institute	Jim Bristow
x	Session Chair	x	Session Chairs	Chair - Alla Lapidus Chair - Mike Fitzgerald
8:45 - 9:30	Keynote	FF0003	DNA Sequencing Technology as the Engine of Scientific Advancement	Dr. Chad Nusbaum
9:30 - 9:50	Speaker 1	FF0044	Large Genome Assembly with 454 Data	Jim Knight
9:50 -10:10	Speaker 2	FF0201	A General Method for Assembling Genomes from Illumina data	Dariusz Przybylski
10:10 - 10:40	Break	x	Beverages and snacks provided	x
10:40 -11:00	Speaker 3	FF0172	Rnnotator: a de novo transcriptome assembly pipeline from short RNA-Seq reads	Jeff Martin
11:00 - 11:20	Speaker 4	FF0006	Phytophthora capsici: genome assembly of a polymorphic organism	Joann Mudge
11:20 - 11:40	Speaker 5	FF0165	Breaking the BLAST bottleneck	Robert Edgar
11:40 - 12:00	Speaker 6	FF0061	Towards experimental gene annotation by deep sequencing	Andrew Bradbury
12:00 - 1:30pm	Lunch	illumina	New Mexican Lunch Buffet Green salad with vegetable garnishes and assorted dressings, Beef fajita quesadilla garnished with lettuce, tomato, sour cream and guacamole, Cheese enchiladas with red chile, Black beans and rice, Beverages	Sponsored by illumina
x	Session Chair	x	Session Chairs	Chair - Donna Muzny Chair - Johar Ali
1:30 - 1:50	Speaker 7	FF0116	The Medicago truncatula HapMap Project: Resequencing of 400 Medicago Genotypes and 48 Sinorhizobium strains using Illumina's 2 nd Generation solexa Technology	Arvind Bharti
1:50 - 2:10	Speaker 8	FF0111	Sequencing the unrearranged human immunoglobulin heavy chain locus from hydatidiform mole	Rene Warren
2:10 - 2:30	Speaker 9	FF0098	Implementation of a cost-effective pyrosequencing process through automation, molecular barcoding and process improvement	Niall Lennon
2:30 - 3:30pm	Genome Finishing (Selected Posters) (10 min each) with question panel	FF0040	Univ. of Florida - Use of optical mapping in bacterial genome finishing	Dibyendu Kumar
		FF0105	Baylor Finishing - Microbe Assembly Strategies and Finishing at BCM-HGSC	Christian Buhay
		FF0100	Sanger Finishing - Gold Standard: Finishing Using Next Generation Technologies	Andries Van Tonder
		FF0127	Broad Finishing - Multiplex Template Preparation for Finishing With Next Generation Sequencing	Anna Montmayeur
		FF0119	LANL Finishing - A high-throughput next generation genome finishing pipeline	Karen Davenport
3:30 - 3:50	Break	x	Beverages and snacks provided	х
3:50 - 5:20pm	Tech Time Talks (15 min each)	FF0209	Caliper - Advanced Nucleic Acid Fractionation for Next Generation Sequencing Sample Preparation	Mark Roskey
		FF0090	Covaris - AFA-based DNA Fragmentation: A Critical First Step for the Optimization of Next Gen Sequencing	Jim Laugharn
		FF0164	OpGen - Use of Optical Mapping to Aid in the Assembly and Finishing of Human Microbiome Genome Projects	Trevor Wagner
		FF0035	Geospiza - Increasing the Scale of Deep Sequencing Data Analysis with BioHDF	Todd Smith
		FF0193	CLC bio - Exploring highly specialized Next Generation Sequencing Analysis tools in CLCbio's Genomics Workbench	Saul Kravitz
		FF0059	NABsys - Hybridization Assisted Nanopore Sequencing	John Oliver
5:45 - 7:45	Happy Hour	LifeTech	Happy Hour at Cowgirls Cafe - Sponsored by Life Technologies (3 drink tickets per person)	Map Will be Provided
7:45 - bedtime	on your own	x	Dinner and night on your own - enjoy	x

Speaker Presentations (June 3rd)

Abstracts are in order of presentation according to Agenda

FF0003

Keynote

Chad Nusbaum

Broad Institute of MIT and Harvard, 320 Charles St. Cambridge MA 02141

DNA Sequencing Technology as the Engine of Scientific Advancement

DNA sequencing technology continues to advance rapidly. By analogy to the computer industry, where general purpose computing is now the rule, DNA sequencing has gone from being a specialized tool to a general purpose one. In particular, the declining cost of generating sequence data provides opportunities both for existing applications to be carried out at greatly increased scale and for new applications to be established that had not been previously affordable. In addition to improvements to sequencing hardware and chemistry, major factors to improve utility of sequencing data include optimization of supporting laboratory processes, invention of new methods for sample preparation, and improvements to algorithms to make use of the data.

In this presentation, I will describe the state of the field of DNA sequencing technology, and will discuss instructive examples elucidating the enabling impact of the changing technology on the pursuit of scientific goals.

NOTES

Large Genome Assembly with 454 Data

James Knight

Roche Diagnostics - 454, Branford, CT 06405, USA

The past year has seen a wide range of de novo projects use 454 reads to sequence large genomes, including mammalian, plant, fish and insect genomes. This talk will highlight several of the projects that have reached the draft assembly stage or better, describe recent improvements to the Newbler assembler for assembling large genomes, and discuss "lessons learned" so far in starting, sequencing and assembling large genome datasets

A General Method for Assembling Genomes from Illumina data

<u>Dariusz Przybylski</u>, Iain A. MacCallum, Sante Gnerre, Joshua Burton, Filipe Ribeiro, Bruce Walker, Ted Sharpe, Giles Hall, Genome Sequencing Platform, Carsten Russ, Chad Nusbaum, David B. Jaffe

Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA

Over the past decade, DNA sequencing costs have dropped about 10,000-fold. The lowest-cost reads from current technologies are short, have a high error rate, and land unevenly on the genome. They are ideally suited to 'resequencing' applications in which a preexisting reference sequence is available; but for *de novo* genome assembly, they are challenging to work with, and results have generally been inferior to those obtained from the old (Sanger method) technology.

Here we demonstrate a practical and general laboratory/computational method for generating high-quality *de novo* assemblies of genomes at the lowest possible cost. Our method starts with 100-base Illumina paired reads from two libraries: one from fragments of size 180 bp (slightly less than twice the read length), and one from fragments of size 3000 bp, via a 'jumping' construction. These two libraries use off-the-shelf methods and provide power that could not be obtained from a single library. We also demonstrate experimental methods for jumping longer fragments to yield a third library, providing even greater potential for long-range connectivity.

We assembled these data using our new version of the ALLPATHS algorithm. This algorithm has been scaled up to work on large genomes and made robust to idiosyncrasies in library construction, variation in coverage, and run-to-run variability in sequence quality, all of which have been critical problems for genome assembly.

We tested our method using a suite of 16 genomes, 9 for which a reference sequence was available and 7 from completely new samples. These genomes range in size from 2 Mb to 2.6 Gb, and in GC content from 19% to 71%. Using the preexisting reference sequences, we assess the completeness, continuity, and accuracy of these assemblies, finding that for the smaller genomes their quality exceeds the general quality level of draft assemblies that had been achieved using Sanger sequencing. For the largest (bushbaby), we obtained contigs and scaffolds having N50 sizes 18 kb and 3.9 Mb, respectively. We compared to a preexisting 3.5x Sanger assembly, and by so doing estimated the *combined* misjoin rate in the two assemblies: one per 10 Mb.

Rnnotator: a de novo transcriptome assembly pipeline from short RNA-Seq reads

<u>Jeff Martin</u>¹, Rob Egan¹, Zide Fang², Vincent Bruno³, Xiandong Meng¹, Matthew Blow¹, Michael Snyder⁴, Zhong Wang¹

- 1. DOE Joint Genome Institute (JGI), Walnut Creek, CA 94598, USA
- 2. School of Public Health, LSU-Health Sciences Center, New Orleans, LA 70112, USA
- 3. Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA
- 4. Department of Genetics, Stanford University Medical School, Stanford, CA 94305-5120, USA.

RNA-Seq, or high throughput next generation sequencing of RNA, is a power tool for transcriptomics study, however, its application to many organisms has been limited where the genome has not yet been (fully) sequenced. Here, we report a software pipeline, called Rnnotator, that *de novo* assembles transcriptomes exclusively from short read sequences to facilitate functional annotation and expression profiling. We used several transcriptomes of different complexity to evaluate the performance of Rnnotator by assessing the accuracy, completeness, and contiguity (ACC) of the assembled transcripts. From these results we conclude that it is possible to reconstruct full-length transcripts using RNA-Seq data without a reference genome. In addition, we discovered many novel transcribed regions that are absent even from well sequenced genomes, suggesting Rnnotator serves as a complementary approach to analysis based on a reference genome for comprehensive functional genomics.

Phytophthora capsici: genome assembly of a polymorphic organism

<u>Joann Mudge</u>¹, Kurt H. Lamour², Jeremy Schmutz³, Sophien Kamoun⁴, Neil Miller¹, Arvind Bharti¹, Paul M. Richardson⁵, Darren Platt⁵, Igor Grigoriev⁵, Alan Kuo⁵, Olga Chertkov⁶, Cliff S. Han⁶, Chris Detter⁶, Greg D. May¹, William D. Beavis⁷, Jason Affourtit⁸, Michael Egholm⁸, Jim Knight⁸, Stephen F. Kingsmore¹

Phythophthora capsici is a devastating oomycete pathogen of vegetable crops. Using 454 titanium and Sanger sequencing technology, we have sequenced and assembled a draft of this ~60 Mb eukaryotic genome. The assembly includes 15X shotgun coverage of 454 Titanium pyrosequencing sequencing averaging 338 nt, 4.5X coverage of 454 Titanium 20kb paired end sequencing averaging 155 nt, and 5X coverage of Sanger 6 or 36 kb paired end sequencing averaging 900 nt. We have assembled approximately 60 Mb into 2,000 scaffolds (scaffold N50 = 669kb). We recovered over 98% of fulllength cDNAs with less than 1% represented in multiple positions. Approximately 3 Mb of repetitive DNA supported only by unanchored repetitive reads was excluded from the assembly to avoid repeat-induced misjoins. In addition, Solexa cDNA libraries from nine life stages have been sequenced to improve gene annotation, identify SNPs, and compare gene expression levels. The broad and expanding host range of this organism emphasizes the genomic adaptability of this organism. To start to understand the natural diversity within P. capsici we have sequenced reduced representation genomic samples (Floragenex's RAD technology and solexa sequencing) from eight field isolates and analyzed the resulting SNPs. The successes and challenges of doing a 454-based de novo genome sequencing project in an oomycete will be discussed.

¹ National Center for Genome Resources, Santa Fe, New Mexico, USA,² The University of Tennessee, Department of Entomology and Plant Pathology, Knoxville, Tennessee, Hudson Alpha Institute for Biotechnology, Huntsville, AL, USA, ⁴Sainsbury Laboratory, Norwich, UK, ⁵USA Department of Energy Joint Genome Institute, Walnut Creek, California, USA, ⁶Los Alamos National Laboratory, Los Alamos, New Mexico, USA, ⁷Iowa State University, Ames, Iowa, USA, ⁸Roche

Breaking the BLAST bottleneck

Robert Edgar

bob@drive5.com

USEARCH is a new algorithm for high-throughput search of large biological sequence databases. USEARCH is two to three orders of magnitude faster than BLAST for some search tasks, enabling rapid classification by database search or by de novo clustering. USEARCH can search a database of 5M proteins (PFAM-A) at a rate 500x faster than BLASTP with comparable sensitivity at sequence identities above 30-35%. In a nucleotide sequence test based on RFAM (0.5M RNAs), USEARCH was ~100x faster than MEGABLAST and substantially more sensitive (96% vs. 73%). Compared to CD-HIT, a widely-used clustering program, USEARCH is more sensitive, producing higher-quality clusters and enabling clustering at lower identities. USEARCH uses substantially less memory and is often one or more orders of magnitude faster than CD-HIT. USEARCH can cluster one hundred million 16S rRNA reads of length ~300 nt in four hours on a laptop, using less than 50 Mb of memory.

Towards experimental gene annotation by deep sequencing

S. D'Angelo^{1*}, N. Velappan^{1*}, T. Gaiotto¹, D. Sblattero², F. Mignone³ & A.R.M. Bradbury¹

Ideally gene annotation would be carried out by experimental analysis of gene products. However, producing proteins from cloned genes and testing them for specific activities, requires a completely different skill set to that involved in high throughput sequencing. Whereas genomes are sequenced rapidly and cheaply, proteins are produced relatively slowly and at far higher cost: there is an immense mismatch between the requirements and throughput of genome sequencing and experimental gene annotation. There is a need therefore, for a method to analyze protein function that can operate at the high throughput capabilities of sequencing centers.

Although proteins have many different functions, molecular recognition, often mediated by modular domains, is a frequent underlying requirement. Molecular recognition is the most easily experimentally assessed activity that can be directly translated into gene annotation. We have found that when DNA fragments are fused to a ß lactamase folding reporter, fragments of real genes are filtered out, as opposed to spurious ORFs with no biological function [1]. As an example, a fragmented cDNA library filtered using the ß lactamase folding reporter and analyzed using 454 sequencing consisted of over 99% open reading frames, of which over 85% were derived from real genes, as opposed to spurious ORFs [2]. By coupling ORF filtering with phage display, specific binding domains have been selected from such ORF libraries [2, 3]. Within the context of GTL Advanced Annotation, we have applied this technique to the C. thermocellum genome, and carried out selection for binding to cellulose. Preliminary experiments, analyzing small numbers of clones, indicate that of five different positive gene fragments identified, two were derived from known cellulose binding domains, one was a sporulation protein, one was part of ruberythin, and one was a conserved hypothetical protein. These results indicate the power of this technique to identify protein domains of known binding function directly from genomic libraries. In previous experiments [2], we showed that 454 sequencing analysis was far more powerful than the traditional picking of random clones and testing by enzymatic immunosorbant assays (ELISA) for the identification of domains with known specific binding functions. In fact, we calculated that over 500 clones would have to be screened by ELISA in order to find the five commonest genes identified by 454 with a 95% probability.

These results indicate a path forwards for a new high throughput experimental gene annotation paradigm that would exchange protein expression and analysis with DNA based techniques used in high throughput sequencing. Such an annotation pipeline would take the following form: 1) creation of filtered phage display genomic domain libraries from sequenced genomes; 2) selection on many different substrates (e.g. substrates, proteins, metabolites) for each filtered genome library; 3) sequencing of all selection outputs using bar-coded 454 and 4) identification of binding domains by sequence analysis for each selection substrate. This genome neutral technology will be applicable to any intronless genome, and will avoid the need for extensive analysis, primer synthesis or protein expression. While the present idea consists of identifying binding domains, it will be possible to extend the concept to include so-called "activity based probes", that would allow the same pipeline to include annotation for enzymatic activity as well.

- 1. Zacchi, P., et al., Selecting open reading frames from DNA. Genome Res, 2003. 13(5): p. 980-90.
- 2. Di Niro, R., et al., Rapid interactome profiling by massive sequencing. Nucleic Acids Res, 2010 DOI 10.1093/nar/gkq052.
- 3. Di Niro, R., et al., Characterizing monoclonal antibody epitopes by filtered gene fragment phage display. Biochem J, 2005. **388**: p. 889–894.

¹Los Alamos National Laboratory, Los Alamos, NM, USA, ²University of Eastern Piedmont, Novara, Italy, ³University of Milan, Milan, Italy, *equal contributions

Lunch

12:00 - 1:30pm

Sponsored by



Notes

The *Medicago truncatula* HapMap Project: Resequencing of 400 *Medicago* Genotypes and 48 *Sinorhizobium* strains using Illumina's 2nd Generation solexa Technology

<u>Arvind K. Bharti</u>¹, Antoine Branca², Tim Paape², Roman Briskine³, Peng Zhou³, Shelley Wang³, Roxanne Denny³, Andrew Farmer¹, John A. Crow¹, Jimmy E. Woodward¹, Joann Mudge¹, Gregory D. May¹, Peter L. Tiffin², Nevin D. Young³

The genome of the model legume *Medicago truncatula* Jemalong A17 has been sequenced by an International consortium using a BAC-by-BAC approach. Legumes are the only plant group capable of biological nitrogen fixation in symbiosis with soil bacteria. As a platform for exploring the genetics of this mutualist interaction with Sinorhizobium, we are resequencing ~400 genotypes of Medicago using Illumina's 2nd generation solexa technology. 30 lines have been deeply resequenced (24x-34x) while the remainder are being sequenced at ~5x coverage. These reads have been aligned to the reference A17 genome using GSNAP after optimizing parameters. These alignments form the basis for defining variation including SNPs (single nucleotide polymorphisms), INDELs (insertion/deletions) and CNVs (copy number variants) using our in-house variant detection pipeline (Alpheus). The initial variant calls will be augmented with imputed SNPs, genome segments with shared ancestry (haplotypes) identified, and population structure elucidated thereby create a long-term, communityaccessible genome-scale association mapping resource. Variant reports for 3 genotypes are already available through www.medicagohapmap.org and underlying reads have been deposited in NCBI's Short Read Archive.

Variants detected with solexa sequencing have been validated by comparing to variants sequenced by the classic Sanger method. Sanger-sequenced regions were chosen randomly throughout the genome among exons, introns and intergenic space. We have begun to generate preliminary patterns of polymorphism in Medicago across the lines completed so far.

Moreover we have also carried out whole genome solexa shotgun sequencing (100x-200x) and *de novo* assembly (ABySS) of 48 bacterial genotypes (*Sinorhizobium meliloti* and *S. medicae*). In parallel, greenhouse experiments on the *Medicago* lines involving these diverse bacterial strains are being monitored in order to define the interplay of phenotype-genotype interaction in symbiotic success.

¹National Center for Genome Resources, Santa Fe, NM 87505, USA

²Department of Plant Biology, University of Minnesota, St. Paul, MN 55108, USA

³Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108, USA

Sequencing the unrearranged human immunoglobulin heavy chain locus from hydatidiform mole

Warren RL1, Watson C2, Schein J1, Zeng T1, Breden F2, Holt RA1

1BC Cancer Agency, Genome Sciences Centre, 675 West 10th Avenue, Vancouver, BC V5Z 1L3, 2Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6

Diversity in antibody binding specificity is generated largely by somatic rearrangement of heavy chain (IgH) and light chain (IgL and IgK) gene segments in B cells. This mechanism is key to adaptive immunity. In addition, immune loci carry germline variation that can confer susceptibility to disease [1-3]. In NCBI_hg19 the 1.28 Mbp IgH locus is represented by a consensus sequence [4] that is a mosaic derived from diploid libraries prepared from lymphoblastoid cell lines [5] and peripheral blood cells that would be expected to include Ig-rearranged B cells [6]. Ubiquitous interspersed repeats,

background somatic recombination, and the high sequence similarity between V(D)J gene segments make assembly of immune loci such as IgH a particularly difficult task. An independent, high-quality IgH reference sequence would be an important resource for the immunogenetics community.

Here, we describe initial sequencing and finishing efforts of the IgH locus using BACs clones from a library constructed from a human hydatidiform mole cell line [7]. A hydatidiform mole is a pregnancy abnormality generated by fertilization of an empty egg, thus the cell line provides haploid genomic DNA from a non-lymphoid source. Overlapping fingerprinted BACs that span the IgH locus were chosen from the mole library, sheared, barcoded, pooled and sequenced to produce 26.7M 76nt paired Illumina reads, giving ~1950x average coverage. Qualitytrimmed reads were assembled *de novo* using our SSAKE v3.4.1 assembler [8], which can handle unequally covered regions. The initial assembly yielded 1,293 contigs (N50=2.5kbp, MAX=19kbp) and 689 scaffolds (N50=9.8kbp, MAX=53kbp). It is being further refined, and gaps are being closed by light coverage of a 5kb plasmid library and primer walking. Once the IgH reference sequence is completed we intend to apply this same approach to the other immune receptor loci.

- 1. Feeney AJ et al. 1996. J Clin Invest. 97:2277-2282
- 2. Cho ML et al. 2003. Lupus. 12:400-405
- 3. Hallmayer J et al. 2009. Nat Genet. 41:708-711
- 4. Matsuda F et al. 1998. J Exp Med. 188:2151-2162
- 5. Larin Z et al. 1991. Proc Natl Acad Sci U S A. 88:4123-4127
- 6. Kodaira M et al. 1986. J Mol Biol. 190:529-541
- 7. Fan JB et al. 2002. Genomics 79:58-62.
- 8. Warren RL et al. 2007. Bioinformatics 23:500-501.

Implementation of a cost-effective pyrosequencing process through automation, molecular barcoding and process improvement

Niall J. Lennon1, Robert E. Lintner1, Scott Anderson1, Pablo Alvarez2, Andrew Barry1, William Brockman3, Riza Daza1, Rachel Erlich1, Georgia Giannoukos1, Lisa Green1, Andrew Hollinger1, Cindi A. Hoover4, David B. Jaffe1, Frank Juhn1, Danielle McCarthy1, Danielle Perrin1, Karen Ponchner1, Taryn Powers1, Kamran Rizzolo1, Dana Robbins1, Elizabeth Ryan1, Carsten Russ1, Todd Sparrow1, John Stalker1, Scott Steelman1, Michael Weiand1, Andrew Zimmer1, Matthew R. Henn1, Chad Nusbaum1 and Robert Nicol1

- 1 Broad Institute of MIT & Harvard, 320 Charles St., Cambridge, MA 02141, USA
- 2 Akamai Technologies Inc., Cambridge, MA 02142, USA
- 3 Google Inc., Cambridge, MA 02142, USA
- 4 Joint Genome Institute, Walnut Creek, CA 94598, USA

The combination of long reads and short run times make the Roche 454 Genome Sequencer FLX Titanium platform ideal for *de novo* assembly of small genomes, metagenomic studies and population diversity profiling. However, the standard protocols for sample preparation are labor-intensive, error-prone and not easily scalable. To overcome these limitations, we have built an automated, fully scalable process for library construction in support of 454 technology. All steps are carried out by liquid handling automation, enabling large numbers of sequence-ready libraries to be constructed in parallel. Sample handling errors and cross-contamination are minimized through the use of end-to-end barcoding of tubes and microtiter plates, along with molecular barcodes in the library constructs themselves. In particular, iron bead-based size selection and cleanup steps have been devised that are fully amenable to automation, eliminating a major bottleneck and significant source of error. Using this methodology an experienced technician can create up to ninety-six sequence-ready 454 libraries in two days, a dramatic improvement over the standard tube-based method.

Notes

Use of optical mapping in bacterial genome finishing

Dibyendu Kumar and William G Farmerie

Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL-32611

The cost-efficiency of modern DNA sequencing technology, such as the Roche 454 GS-FLX, allows individual investigators to undertake bacterial genome projects that were not affordable only a few years ago. Our core laboratory has several ongoing bacterial genome projects presenting a variety of challenges to genome assembly and closure. Several factors contribute to these challenges; including sequence repeats versus read length, intrinsic sequencing errors, and dynamic genome rearrangements. Together these factors complicate genome closure when using shotgun DNA sequencing alone. The genome finisher may experience difficulty validating their assembly in the absence of a physical map. To address this problem, we adopted whole-genome optical mapping as a tool to validate bacterial genome assemblies. OpGen, Inc. (Gaithersburg, Maryland) prepared the optical maps used in this project.

Briefly, an optical map is a complete genome restriction map deduced from a number of partial restriction maps. Optical maps are generated by spreading carefully extracted genomic DNA onto a treated glass surface containing many narrow channels, followed by digestion *in situ* with restriction enzymes. About 50–100 contiguous restriction fragments with sizes approaching up to one-third of the whole genome are selected and optically measured. The overlapping partial optical contigs are combined by alignment software to produce a contiguous whole genome restriction map. The contiguous optical map can be aligned and compared with the *in silico* restriction map determined for the partially complete whole-genome assembly. We successfully used optical mapping for guiding the closure of four closely related bacterial genomes. The optical map allowed us to identify assembly errors not possible using shotgun DNA sequencing data alone. Thus, we conclude that, in order to ensure the accuracy of a finished bacterial genome, optical mapping is an important tool to validate *de-novo* assemblies generated by next-generation DNA sequencing.

Microbe Assembly Strategies and Finishing at BCM-HGSC

<u>Christian J. Buhay</u>, Yuan-Qing Wu, Shannon Dugan-Rocha, Sandra L. Lee, Michael E. Holder, Yan Ding, Xiang Qin, Vandita Joshi, Irene Newsham, Joseph Petrosino, Sarah K. Highlander, Donna M. Muzny and Richard A. Gibbs

Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030

BCM-HGSC is one of four major centers involved in the Human Microbiome Project. To date, 105 microbial genomes have been assembled as part of the jumpstart portion of this project. 37 of those genomes are currently in the finishing pipeline with 89% completed at various finishing levels. These finishing levels range from automated and or/manual work that provides an "improved" quality over the initial draft sequence to high quality or "gold standard" finished genomes with no unresolved or unconfirmed regions.

As we have continued to evaluate additional microbial genomes, we have identified some primary issues related to assembly and finishing. In addition to informatics based pipeline improvements, experiments with larger 8kb – 10kb insert libraries have been conducted to improve assembly accuracy and contiguity. One major hurdle in finishing microbial genomes has been dealing with large repeats such as rRNA operons. These larger insert libraries enable us to span these repeats and provide better assembly contiguity.

This new technique has been applied to additional microbes currently in the finishing pipeline and further comparisons of 3kb long tag paired ends, 3kb recombi paired ends, and 8kb recombi paired ends for 13 genomes have been completed. Sequence coverage was normalized within each organism as well as across all 13 genomes. Eleven additional 8kb recombi paired-end assemblies were also assembled and analyzed. Assemblies of these microbes with 8kb libraries showed great improvement. More significantly, scaffolding statistics showed that for many of these organisms, most of the genome was found in one large scaffold. For one microbe, *Aeromicrobium marinum* DSM 15272, there was an 8-fold increase in the scaffold N50 from 384 kb in the 3kb long tag paired-end and recombi paired-end assemblies to 3 Mb in the 8kb recombi paired-end assembly. Success with this new technique means most of the genome will be ordered and oriented upon initial assembly, generating higher quality draft assemblies that would require significantly less finishing effort. We are continuing to evaluate and optimize this strategy with additional microbes and plan to implement it for all reference strain sequence assemblies and finishing.

Gold Standard: Finishing Using Next Generation Technologies

Andries J. Van Tonder

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA, UK

A pipeline for finishing genomes to published criteria has been designed using *Staphylococcus lugdunensis*. This pipeline is scalable and applicable to a variety of data sets consisting of next generation sequencing technologies.

S. lugdunensis is a Gram-positive bacterium which occurs as a commensal on human skin. The 5Mb genome was sequenced using both the Illumina GAII platform to a depth of 630x coverage and the 454/Roche Titanium FLX3 platform to a depth of 16x coverage. Illumina data was assembled using Velvet before IMAGE (Iterative Mapping and Assembly for Gap Elimination; Tsai In press) improved the assembly by iterative gap closure. Velvet consensus sequences were combined with 454 data to create an assembly using Newbler. The Ace file generated by Newbler could then be converted to be viewable in a sequence editor such as consed or gap4. ABACAS (Algorithmic Based Automatic Contiguation of Assembled Shotgun Sequence; Assefa 2009), along with the 454scaffold fina file from Newbler, identified the correct contig order. This allowed manual finishing, including the ordering of reactions such as PCRs, to be undertaken. Once the genome was contiguous scripts such as iCORN (Iterative Correction of Reference Nucleotides; Otto Unpublished), which provided sequence correction by mapping Illumina reads back to the assembly, were used to improve the quality of the data.

A set of criteria was applied to ensure that the genome was finished to Gold Standard, as detailed by Chain et al. (2009). This requires every position in the sequence be covered by at least two technologies and the error rate of the consensus needs to be less than 1 error per 100,000 base pairs. The assembly was checked to ensure that it was free of misassemblies and that repeat regions had been ordered and correctly assembled.

The described pipeline will be presented detailing the work flow and methodologies.

Multiplex Template Preparation for Finishing With Next Generation Sequencing

<u>Anna Montmayeu</u>r, Gary Gearin, Amr Abouelleil, Annie Lui, Niall Lennon, Rachel Erlich, Harindra M. Arachchi, Chelsea Dunbar, Jennifer Baldwin, Bruce W. Birren and Michael G. FitzGerald

Broad Institute of MIT and Harvard, Cambridge, MA

Finished genome sequence is an important component of our bacterial pathogen research program. Complete sequence facilitates identification of resistance mechanisms, epidemiological work and provides a complete reference for mapping related strains.

Finishing by means of next generation sequencing technologies is only cost effective when a very large number of targets are processed simultaneously. In the case of the Roche-454 GS FLX Titanium sequencer this calls for the labor-intensive steps of large-scale PCR amplicon generation followed by sample barcoding and 454-adapter addition, so that samples can be pooled together prior to sequencing.

In an effort to simplify sample preparation we turned to molecular inversion probes (MIPs), also known as "padlock probes". MIPs were originally used for high-throughput, multiplex SNP detection. Recently they have also been used for the multiplexed capture and sequencing of a large number of small exons.

We are conducting a series of pilot experiments using pathogen genomes currently involved in our active finishing process. Targets are designed around gap regions of the pathogens, *Neisseria gonorrhoeae* MS11, *Enterococcus casseliflavus* 899205 and *Staphyloccus aureus* 55/2053. All assemblies are based on 454-FLX data and are assembled using Newbler. We will present an evaluation of MIP results for 64 gaps. All have reference data available from our production (Sanger) finishing process. Pilot targets range from 0.2kb to 2.5kb. Preliminary results are quite positive and demonstrate our ability to capture and derive sequence from targets up to 1.5kb. This project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900006C.

LANL Finishing: a High-Throughput Finishing Pipeline with Second Generation Sequencing Technologies.

K. Davenport*, L. Goodwin, L. Meincke, O. Chertkov, C. Han, and C. Detter

Los Alamos National Laboratory, Los Alamos, NM.

New developments in high-throughput finishing at LANL include automated start-up, automated closeout, and project management. With the increasing number of projects coming into our finishing pipeline we are finding new ways to streamline the finishing process. We are working to decrease the amount of time a project spends in manual finishing by improving and automating the first cycle of finishing as well as implementing a formal project management system.

.

Genome Finishing Notes

Advanced Nucleic Acid Fractionation for Next Generation Sequencing Sample Preparation

Mark Roskey

Applied Biology and R&D, aliper Life Sciences

Separating nucleic acids by size with subsequent isolation of specific fragments is a fundamental, but tedious, molecular biology technique. As second generation sequencing platforms gain utility, it has become clear that reproducible and scalable sizing and isolation has become a bottle neck for many applications of new sequencing technologies. Caliper Life Sciences has developed and commercialized instruments that utilize microfluidics to achieve rapid and high resolution electrophoretic separations. While we have focused most of our efforts on analysis applications (e.g. DNA, RNA and protein sizing), we have also previously demonstrated preparative applications, such as fractionation of nucleic acid for cloning. We have now developed a commercial solution, the LabChip XT, that will simplify and improve nucleic acid fractionation. By using intersecting microfluidic channels, optical detection and computer control, we can automatically extract a target band during separation and route the selected material to a collection well. Using this process we have successfully isolated narrow bands (±5%) in the range of 100-500 bp from a fragmented DNA sample. The isolated nucleic acid is PCR-ready and the fractionation is complete in 30-60 minutes. The presented material will describe the fundamentals of our microfluidics-based solution, as well as highlight some early data and collaborative work.

Covaris AFA-based DNA Fragmentation: A Critical First Step for the Optimization of Next Gen Sequencing

Jim Laugharn, Guillaume Durin, and Edwin Rudd

Covaris, Inc., 14 Gill Street, Unit H, Woburn, MA 01801 USA

The controlled generation of random fragments from genomic DNA, with high recovery and high quality, is the most critical sample preparation step required by all the currently available next generation sequencing platforms. As the requirements for improved sensitivity and application depth of these platforms increase, high demands are also placed on the standardization of the sample fragmentation process to be easily transferable between laboratories (intra- and inter-lab) and scalable to enable both automation and higher throughput requirements. To be successful in the next-gen sequencing market, it requires utilizing an efficient, easy-to-use, high throughput, and highly reproducible DNA shearing technology. Standardizing on high accuracy and high precision will become a fundamental prerequisite of all nucleic acid fragmenting sample preparation processes, as genomic sequencing technologies continue to develop; particularly in clinical trials and clinical applications.

Although nebulization (wind shear), enzymatic digestions, hydrodynamic shearing, and low frequency sonicators have been used to shear DNA, they all have significant intrinsic disadvantages which make them the weak points in next generation sequencing workflows. These limitations and disadvantages include; thermal and sequence specific biased fragmentation, thermal degradation, valuable sample loss, automation incompatibility, and user-dependent reproducibility issues. In stark contrast, the closed vessel, non-contact, isothermal processing, and easily automatable Covaris Adaptive Focused AcousticsTM (AFA) process stands out as the most widely used and preferred method for easily and reproducibly shearing DNA.

We will show how AFA enables standardization of a sample preparation process and work flow. We will present data which demonstrates the reproducibility of our validated protocols to efficiently fragment DNA in the 100bp to 5kb size range, and we will discuss the use of our technology in chromatin shearing for ChIP-chip, ChIP-Seq applications. Performance data from new products will be presented on a designed—for-automation, pre-assembled 96 microTUBE plate and the parallel LE220 instrument (automation-ready) for high-throughput laboratories.

Use of Optical Mapping to Aid in the Assembly and Finishing of Human Microbiome Genome Projects

T. K. Wagner, E. A. Meudt, E.B. Zentz. OpGen Inc., Gaithersburg, MD 20878.

One of the goals of the Human Microbiome Project (HMP) is to sequence the whole genomes of over 1,000 human commensals. However, currently less than 2% (25/1482) are considered as contiguous and finished. Optical Mapping is a technology that can rapidly yield whole-genome ordered restriction maps that can be used as a whole-genome scaffold to position, order, orient, and validate the assembly of DNA sequencing contigs. The hypothesis of this study was that Optical Mapping could be used to place the majority of contigs 40 kb and larger of the selected HMP reference genomes.

Whole-genome Optical Maps were generated from 7 HMP reference microbes ordered from ATCC, which included *Chryseobacterium gleum* (ATCC 35910), *Citrobacter youngae* (ATCC 29220), *Enterobacter cancerogenus* (ATCC 35316), *Neisseria cinerea* (ATCC 14685), *Neisseria sicca* (ATCC 29256), *Streptococcus infantarius infantarius* (ATCC BAA-102), and *Weissella paramesenteroides* (ATCC 33313). Contigs of wholegenome DNA sequence for each bacteria were obtained from GenBank, imported into MapSolver™ software to generate an *in silico* restriction map of each contig, and compared to the Optical Maps.

Over 85% of the contigs 40 kb and larger were positioned and ordered against the Optical Maps. The placement of contigs resulted in an average of over 75% coverage of the unfinished genomes. The gaps between contigs were located and the sizes determined. Potential misassemblies were identified in the DNA sequence contigs.

These data support the ability of Optical Mapping to position, order, orient, and validate the assembly of DNA sequencing contigs for whole-genome sequencing projects. The use of Optical Mapping could not only help assemble the 98% of the HMP reference genomes that are still considered unfinished but also decrease the time to do so by locating gaps and identifying misassemblies in the DNA sequence contigs.

Increasing the Scale of Deep Sequencing Data Analysis with BioHDF

<u>Todd Smith</u> (1), Christopher E Mason (2), Paul Zumbo (2), Mike Folk (3), Dana Robinson (3), Ruth Aydt (4), Mark Welsh (1), Eric Smith (1), N. Eric Olson (1),

1. Geospiza, Inc. 100 West Harrison N. Tower 330, Seattle WA 98119 2. Department of Physiology and Biophysics, Weil Cornell Medical College, 1305 York Ave., New York NY, 10021 3. The HDF Group, 1901 S. First St., Champaign IL 61820

Next Generation DNA Sequencing (NGS) technologies are powerful tools for rapidly sequencing genomes and studying functional genomics. Presently, the value of NGS technology has been largely demonstrated on individual sample analyses (1-3). The full potential of NGS will be realized when it can be used in multisample experiments that involve different measurements and include replicates, and controls to make valid statistical comparisons. Arguably, improvements in current technology, and soon to be available "third" generation systems, will make it possible to simultaneously measure 100's to1000's of individual samples in single experiments to study transcription, alternative splicing, and how sequences vary between individuals and within expressed genes. However, several bioinformatics systems challenges must be overcome to effectively manage both the volumes of data being produced and the complexity of processing the numerous datasets that will be generated.

Future bioinformatics applications need to be developed on common standard infrastructures that can reduce overall data storage, increase data processing performance, integrate information from multiple sources and are self-describing. HDF technologies meet all of these requirements, have a long history, and are widely used in data-intensive science communities. They consist of general data file formats, software libraries and tools for manipulating the data. Compared to emerging standards such as the SAM/BAM formats, HDF5-based systems demonstrate improved I/O performance and improved methods to reduce data storage. HDF5 is also more extensible and can support multiple data indexes and store multiple data types. For these reasons, HDF5 and its BioHDF implementation are well qualified as standards for implementing data models in binary formats to support the next generation of bioinformatics applications. Through this presentation we will demonstrate BioHDF's latest features in NGS applications that target transcription analysis and resequencing.

Exploring highly specialized Next Generation Sequencing Analysis tools in CLCbio's Genomics Workbench

Saul Kravitz, Dan Conway, and Joe Salvatore

CLC Bio, Cambridge Massachusetts 02142

The Genomics Workbench allows the user to create multiple experiments from it's suite of analysis tools from RNA-Seq, CHip-Seq, and SNP discovery to the ability to de Novo assemble genomes of unlimited size. CLCbio's Genomics Workbench allows the seamless integration of next generation sequencing data in one experiment from all sequencing platforms, which includes 454, Illumina, ABI-SOLiD (color-space) and Helicos.

Hybridization Assisted Nanopore Sequencing

<u>John S. Oliver</u>, Heidi Geiser, Peter lanakiev, Peter Goldstein, Maryam Jouzi, Peter H. Goldstein

NABsys, Inc., Providence, RI 02903

NABsys Inc. is developing a nanopore based DNA sequencing instrument. In the past decade it has been shown that nanopores are sensitive resistive pulse detectors capable of discriminating between individual DNA molecules based on length, state of folding, or duplex versus single stranded forms.

NABsys' technology, termed Hybridization Assisted Nanopore Sequencing (HANS) relaxes the stringent requirement that the nanopore have single base resolution. The nanopore is used to detect hybridization of probes to target DNA and to give relative positional information. The positional accuracy required for assembly of the sequence is reduced by orders of magnitude with respect to direct nanopore sequencing.

We will discuss progress towards implementing HANS using a procedure that allows us to detect probes bound to single-stranded template DNA with high signal-to-noise and high resolution. We will discuss the sample preparation procedure, results of translocation experiments and the ramifications for correct assembly and reconstruction of sequence information.

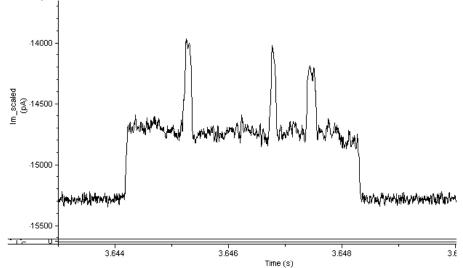


Figure 1: Three probes bound to M13 at positions 1600 bases, 3850 bases, and 4850 bases from the end of the template strand.

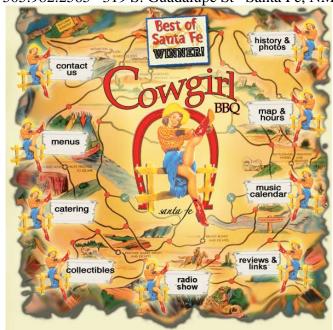
NOTES

NOTES

Happy Hour (x2)

Cowgirls BBQ

505.982.2565 319 S. Guadalupe St Santa Fe, NM



See map on next page!

5:45pm - 7:45pm, June 3rd

3 drink tickets (margaritas, beer, sodas) will be provided per person

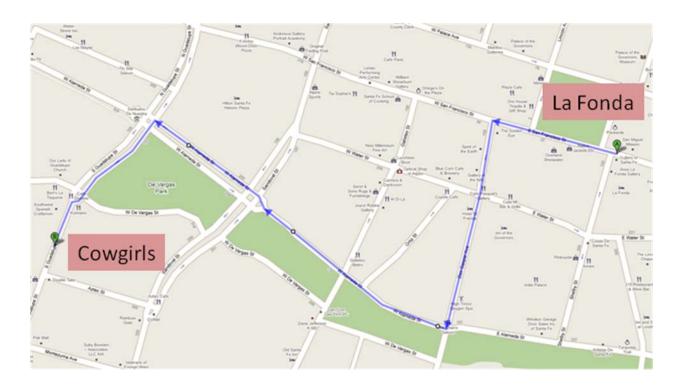
Sponsored by Life Technologies

Enjoy!!!



Map to Cowgirls BBQ

505.982.2565 319 S. Guadalupe St Santa Fe, NM



Total Walking Distance

0.5 miles, 10 minutes

The Legend...

Many years ago, when the cattle roamed free and Cowpokes and Cowgirls rode the range, a sassy young Cowgirl figured out that she could have as much fun smokin' meats and baking fine confections as she could bustin' broncs and rounding up outlaws. So she pulled into the fine bustling city of Santa Fe and noticed that nobody in town was making Barbeque the way she learned out on the range. She built herself a Texas-style barbecue pit and soon enough the sweet and pungent scent of mesquite smoke was wafting down Guadalupe street and within no time at all folks from far and near were lining up for heaping portions of tender mesquite-smoked brisket, ribs and chicken. Never one to sit on her laurels, our intrepid Cowgirl figured out that all those folks chowing down on her now-famous BBQ need something to wash it all down with. Remembering a long-forgotten recipe from the fabled beaches of Mexico, she began making the now-legendary Frozen Margarita and the rest, as we say, is History. Before you could say "Tequila!" the musicians were out playing on the Cowgirl Patio and the party was in full swing.

06/04/2010 - Friday				
Time	Туре	Abstract #	Title	Speaker
7:30 - 8:30am	Breakfast	v	Breakfast Buffet Harvey House French toast with warm maple syrup and butter, Sausage links, Freshly brewed regular and decaffeinated coffee	x
8:30 - 8:45	Intro	FF0232	Welcome Back - Review Genome Standards Paper What's next?	Patrick Chain
x	Session Chair	x	Session Chairs	Chair - Patrick Chain Chair - Darren Grafham
8:45 - 9:30	Keynote	FF0048	Analysis of metagenomic sequences, from megabases to terabases	Dr. Folker Meyer
9:30 – 9:50	Speaker 1	FF0142	Community genomic analysis of relatively low diversity ecosystems: opportunities and challenges for up-scaling	Jill Banfield
9:50 – 10:10	Speaker 2	FF0115	Whole genome shotgun metagenomic sequencing, assembly, and analysis of human microbial communities associated with health and disease	Joe Petrosino
10:10 – 10:30	Break	x	Beverages and snacks provided	x
10:30 – 10:50	Speaker 3	FF0218	Grepping Life: A New Paradigm for Analyzing Metagenomic Data	Joel Berendzen
10:50 – 11:10	Speaker 4	FF0160	One Microbial Cell, One Complete Genome	Tanja Woyke
11:10 – 11:30	Speaker 5	FF0135	Single Amplified Whole Genome Shotgun – Applications	Eric Pelletier
11:30 – 11:50	Speaker 6	FF0208	Whole Genome Sequencing of Individual Bacterial Cells from the Ocean	Rex Malmstrom
11:50 – 12:10	Speaker 7	FF0055	DNA sequencing from single bacterial cells using the multiple displacement amplification (MDA) reaction	Roger Lasken
12:10 - 12:30	Closing Discussions	x	Closing Discussions - discuss next year's meeting	Chair - Chris Detter
12:30 - 2:00pm	Lunch & Close of meeting		La Fiesta Plaza Lunch Pork tenderloin served with papaya coulis and mango salsa Fresh seasonal vegetables, Chef's choice of starch, Sourdough rolls and butter, Beverages	x

г

Speaker Presentations (June 4th)

Abstracts are in order of presentation according to Agenda

Patrick Chain Review of Genome Standards Paper – What's Next?

FF0232

Genome Project Standards in a New Era of Sequencing

P. S. G. Chain, 123*†§ D. V. Grafham, 4†§ R. S. Fulton, 5† M. G. FitzGerald, 6† J. Hostetler, 7† D. Muzny, 8† J. Ali, 9 B. Birren, 6 D. C. Bruce, 110 C. Buhay, 8 J. R. Cole, 3 Y. Ding, 8 S. Dugan, 8 D. Field,11 G. M. Garrity, 3 R. Gibbs, 8 T. Graves, 5 C. S. Han, 110 S. H. Harrison, 3* S. Highlander, 8 P. Hugenholtz, 1 H. M. Khouri, 12 C. D. Kodira, 6* E. Kolker, 1314 N. C. Kyrpides, 1 D. Lang, 12 A. Lapidus, 1 S. A. Malfatti, 12 V. Markowitz, 15 T. Metha, 6 K. E. Nelson, 7 J. Parkhill, 4 S. Pitluck, 1 X. Qin, 8 T. D. Read, 16 J. Schmutz, 17 S. Sozhamannan, 18 P. Sterk, 11 R. L. Strausberg, 7 G. Sutton, 7 N. R. Thomson, 4 J. M. Tiedje, 3 G. Weinstock, 5 A. Wollam, 5 Genomic Standards Consortium Human Microbiome Project Jumpstart Consortium J. C. Detter10†‡‡

1U.S. Department of Energy Joint Genome Institute. 2Lawrence Livermore National Laboratory. 3Michigan State University. 4The Sanger Institute. 5Washington University School of Medicine. 6The Broad Institute. 7J. Craig Venter Institute. 8Baylor College of Medicine. 9Ontario Institute for Cancer Research. 10Los Alamos National Laboratory. 11Natural Environmental Research Council Centre for Ecology and Hydrology. 12National Center for Biotechnology Information. 13Seattle Children's Hospital and Research Institute. 14University of Washington School of Medicine. 15Lawrence Berkeley National Laboratory. 16Emory GRA (Georgia Research Alliance) Genomics Center. 17HudsonAlpha Institute. 18Naval Medical Research Center. *Full affi liations are available on *Science* Online. †Finishing in the Future Working Group members. ‡These authors contributed equally to organizing this work. §Authors for correspondence. E-mail: pchain@lanl.gov (P.S.G.C.); dg1@sanger.ac.uk (D.V.G.)

www.sciencemag.org SCIENCE VOL 326 9 OCTOBER 2009

NOTES

Keynote

Folker Meyer
Argonne National Laboratory

Analysis of metagenomic sequences, from megabases to terabases

NOTES

Community genomic analysis of relatively low diversity ecosystems: opportunities and challenges for up-scaling

Brian Thomas, Christine Sun, Chris Miller, and Jill Banfield*

University of California Berkeley

A focus of our laboratory is the genomics-based analysis of natural microbial communities from systems with sufficient dominance by a few populations to enable extensive reconstruction of population genomic datasets. Environments such as acid mine drainage, hypersaline lakes, subsurface bioremediation, and certain human microbiomes have proven to be particularly amenable to analysis via this approach. In addition to yielding near-complete genomes for organisms that comprise as little as ~1% of communities, along with their associated viruses and plasmids, the product of such analyses is a snapshot of the variation that distinguishes individuals. This variation has been analyzed to provide insight into evolutionary processes and microbial community dynamics. Of particular interest have been studies of time series samples that reveal changing bacterial and archaeal immunity to mobile elements and co-evolution viruses. The research has motivated the development of some tools. However, many challenges relating to automated analyses of variation in more complex systems and integration of different data types remain.

Whole genome shotgun metagenomic sequencing, assembly, and analysis of human microbial communities associated with health and disease

^{1,2}Joseph F. Petrosino, ²Xiang Qin, ²Christian Buhay, ²Michael Holder, ²Kim C. Worley, ²Donna M. Muzny, ^{1,2}Sarah K. Highlander and ²Richard A. Gibbs

¹Department of Molecular Virology and Microbiology, ²Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

There are greater than 10 times the number of bacteria in the GI tract than cells in the human body, and these organisms encode over 100 times the number of genes than in the human genome. We are using Next Generation (NexGen) sequencing technologies to determine the structure of these microbial communities in healthy individuals and in those with infectious or hereditary diseases such as Diabetes, Leukemia, HIV and Travelers' Diarrhea.

NexGen amplicon strategies enable deep sequencing of 16S rDNA bacterial gene regions, while whole genome shotgun (WGS) metagenomic strategies are used to identify the total genomic content of a given sample without the biases of 16S gene amplification. WGS sequencing is a powerful strategy for metagenomic studies with several advantages that enable detailed studies of relationships between individual organisms and the pangenome of microbial communities. We have been developing tools to generate and analyze WGS metagenomic sequence data and are evaluating: the quality and utility of these data sets, the ability to use very small amounts of DNA for library construction (approaching 10ng of template), and how WGS analyses can be streamlined and improved. In addition, we have been assessing different ways to assemble WGS metagenomic data using various assembly tools and mixed platform data sets, and we are evaluating the utility of these assemblies for determining organism and genetic content. Being able to compare the microbiota in our healthy and diseased populations at the pangenomic level will further strengthen the impact of our microbiome disease association projects and will facilitate answering the question as to how these communities impact human health.

Grepping Life: A New Paradigm for Analyzing Metagenomic Data

Joel Berendzen

Los Alamos National Laboratory, Los Alamos, NM

Analysis methods based on local patterns offer enormous computational and conceptual advantages over methods based on global-similarity measures such as BLAST. We set out to revisit phylogenetic methods with a mantra of "What Would Google Do?". We based our method around signatures of amino-acid 10-mers: signatures because they bypass paralog problems, amino-acid because they persist across larger phylogenetic distances than do nucleic acids, 10 because then matches are highly significant, and k-mers because solid patterns can be quickly organized and searched via index structures. Some 20 million orthogenomic amino-acid 10-mer signatures are sufficient to cover all of current sequenced bacterial diversity at the phylum level while allowing detection of reads from never-before-sequenced phyla with reasonable likelihood.

Our method uses the occurrence of these orthogenomic signatures in 6-frame-translated metagenomic reads to associate individual reads with phylogeny and function. We also identify 10-mers that appear with unusually high frequencies as metagenomic signatures of that community. These techniques enable quantitative exploration of microbial ecology without reliance upon directed primers, gene calling, genome assembly, or operational taxonomic units.

Computational demands of our method could be be met with a laptop for small projects or scaled out with map-reduce to large enterprises. Sequencing demands are modest compared with methods which benefit from assembly; community composition and complexity can be estimated from data sets of a few megabases. Reads from large eukaryotic genomes can be tolerated or incorporated in the analysis. We sketch some novel applications our approach enables.

One Microbial Cell, One Complete Genome

<u>Tanja Woyke¹</u>, Alicia Clum¹, Alex Copeland¹, Konstantinos Mavromatis¹, Damon Tighe¹, Wendy Schackwitz¹, Alla Lapidus¹, Dongying Wu¹, Janet Lee¹, John McCutcheon², Bradon McDonald², Nancy Moran², Jim Bristow¹ and Jan-Fang Cheng¹

While the bulk of the finished microbial genomes sequenced to date are derived from cultured bacterial and archaeal representatives, the vast majority of microorganisms elude current culturing attempts, severely limiting the ability to recover complete or even partial genomes from these environmental species. Single cell genomics is a novel culture-independent approach, which enables access to the genetic material of an individual cell. No single cell genome has to our knowledge been closed and finished to date. Here we report the completed genome from an uncultured single cell of Candidatus Sulcia muelleri DMIN. Digital PCR on single symbiont cells isolated from the bacteriome of the green sharpshooter Draeculacephala minerva bacteriome allowed us to assess that this bacteria is polyploid with genome copies ranging from approximately 200-900 per cell, making it a most suitable target for single cell finishing efforts. For single cell shotgun sequencing, an individual Sulcia cell was isolated and whole genome amplified by multiple displacement amplification (MDA). Sanger-based finishing methods allowed us to close the genome. To verify the correctness of our single cell genome and exclude MDA-derived artifacts, we independently shotgun sequenced and assembled the Sulcia genome from pooled bacteriomes using a metagenomic approach, yielding a nearly identical genome. Four variations we detected appear to be genuine biological differences between the two samples. Comparison of the single cell genome with bacteriome metagenomic sequence data detected two single nucleotide polymorphisms (SNPs), indicating extremely low genetic diversity within a Sulcia population. This study demonstrates the power of single cell genomics to generate a complete, high quality, non-composite reference genome within an environmental sample, which can be used for population genetic analyzes.

¹ DOE Joint Genome Institute, Walnut Creek, California 94598, United States of America., ² University of Arizona, Tucson, Arizona 85721, United States of America.

Single Amplified Whole Genome Shotgun – Applications

Adriana Alberti 1, <u>Eric Pelletier 1,2,3</u>, Christophe Bataille 1, Magali Boutard 1, Sophie Oztas 1, Arnaud Couloux 1, Patrick Wincker 1,2,3 & Valerie Barbe 1

1.Genoscope – CEA, 2.CNRS UMR-8030, 3.Université Évry-Val d'Essonne, 2 rue Gaston Crémieux, 91000 Évry, France

New sequencing technology developments, along with genomic amplification techniques,, now allows to have satisfactory access to genomic information of a single cell at a large scale. Such approaches are now used at Genoscope to explore natural environments, such as bacterial communities, endocellular symbionts or monocelullar eukaryotes, giving a credible alternative to classical metagenomic approaches.

An overview of actually ongoing projects and results enlightening these aspects will be given.

Whole Genome Sequencing of Individual Bacterial Cells from the Ocean

Rex Malmstrom

Lawrence Berkeley National Laboratory

Genome sequencing is one of the most powerful tools for exploring the metabolic potential of microorganisms, but our inability to culture the vast majority of these microbes presents a serious bottleneck to the exploration of microbial diversity. Advances in single-cell whole genome amplification (WGA) and next generation sequencing technologies are now allowing us to bypass this bottleneck. Assembling whole genomes from individual cells, however, presents many challenges including the co-amplification of contaminating DNA, the formation of genome rearrangements, and huge variation in coverage along the genome. These challenges were explored systematically with a previously sequenced strain of Prochlorococcus, a unicellular marine cyanobacterium, using a high-throughput pipeline for single-cell WGA that combines fluorescence-activated cell sorting and semi-automated liquid handling. We demonstrated that by using strict handling procedures the impact of contaminating DNA can be virtually eliminated from sequencing libraries while still maintaining throughput that allows for WGA on hundreds of cells at a time. We also confirmed that reads joining non-contiguous genomic regions, which result from rearrangements during WGA, are found ~1 per 10kb, but the potentially deleterious impact of these chimeric reads on de novo assembly can largely be overcome by deep sequencing. Finally, we developed a method that mitigates the huge variations in genome coverage and increases sequencing efficiency. Using our pipeline, 95% of the genome of a single Prochlorococcus cell was assembled de novo into several hundred contigs. Insights from this work provided the tools to confidently explore uncultured clades of Prochlorococcus using single-cell genomics.

DNA sequencing from single bacterial cells using the multiple displacement amplification (MDA) reaction

Roger S. Lasken

J. Craig Venter Institute

Recently developed methods are enabling single-cell sequencing of uncultured microbes. Previously, DNA sequencing of most novel species could only be carried out on bulk DNA extractions from heterogeneous populations. However, a method called multiple displacement amplification (MDA) can now be used to sequence novel microbes from a single cell isolated from the environment with no need to culture the cell. MDA yields micrograms of amplified genomic DNA (Reviewed in T. Ishoey, et al. (2008) Current Opinion in Microbiology, vol. 11, 198-204). About 50-90% of the genome is represented in a typical amplification. Special considerations will be reviewed for sequencing and assembly from amplified DNA. The performance of MDA-generated template is evaluated for 454 and Illumina sequencing methods. De novo assembly for novel microbes will be discussed for Illumina sequencing. The combined use of single cell and metagenomic sequences is also investigated.

NOTES

Discussion Notes

Discussion Notes

2010 SFAF Attendee List					
FF#	Name	Affiliation	email		
1	Chris Detter	Los Alamos National Laboratory - JGI	cdetter@lanl.gov		
2	Isaac Meek	Caliper Life Sciences	Isaac.Meek@caliperls.com		
3	Chad Nusbaum	Broad Institute/MIT	chad@broadinstitute.org		
4	Stephen Kingsmore	National Center for Genome Resources (NCGR)	sfk@ncgr.org		
5	Greg May	National Center for Genome Resources (NCGR)	gdm@ncgr.org		
6	Joann Mudge	National Center for Genome Resources (NCGR)	jm@ncgr.org		
7	X	X	X		
9	Teri Mueller	Roche Diagnostics National Center for Genome Resources (NCGR)	teri.mueller@roche.com		
9 10	Ernie Retzel Frank Tansley	Life Technologies	efr@ncgr.org Frank.Tansley@lifetech.com		
11	Janine Detter	Los Alamos National Laboratory	Janined@lanl.gov		
12	Ron Walters	Pacific Northwest National Laboratory	ra.walters@pnl.gov		
13	Tara Bennink	Edge BioSystems	TBennink@edgebio.com		
14	Holly Coleman	The National Institutes of Health Intramural Sequencing Center (NISC)	hcoleman@mail.nih.gov		
15	Robert Blakesley	NIH Intramural Sequencing Center (NISC)	rblakesl@nhgri.nih.gov		
16	Shelise Brooks	NIH-NHGRI-NISC	sbrooks@mail.nih.gov		
17	Kenneth Frey	Airforce Research Laboratory	Kenneth.Frey@WPAFB.AF.MIL		
18	Fiona Stewart	New England Biolabs	stewart@neb.com		
19	Tim Hunkapiller	Discovery Bio	tim@discoverybio.com		
20	Lee Poeppelman	Airforce Research Laboratory	Lee.Poeppelman@WPAFB.AF.MIL		
21	Armand Dichosa	Los Alamos National Laboratory	armand@lanl.gov		
22	х	Х	Х		
23	Stan Letovsky	Independent	letovsky@bu.edu		
24	Rafal Woycicki	Warsaw University of Life Sciences - SGGW	rafalwoycicki@gmail.com		
25	Michael Rhodes	Life Technologies	Michael.Rhodes@lifetech.com		
26	Marcella Putman	Life Technologies	Marcella.Putman@lifetech.com		
27	Eric Mathur	Synthetic Genomics, Inc.	EMathur@SyntheticGenomics.com		
28	Clotilde Teiling	Roche Applied Science	clotilde.teiling@roche.com		
29	Alistar Bishop	Chemical and Biological Detection Group, Defence Science and Technology Laboratory (DSTL) Porton Down	DWCLEARY@mail.dstl.gov.uk		
30	Claire Fraser-Liggett	Institute for Genome Sciences, University of Maryland School of Medicine	cmfraser@som.umaryland.edu; BAli@som.umaryland.edu		
31	Michael Fitzsimons	Los Alamos National Laboratory	msfitz@lanl.gov		
32	Ken Taylor	Integrated DNA Technologies	ktaylor@idtdna.com		
33	Alfredo Lopez De Leon	Novozymes, Inc.	ALLO@novozymes.com		
34	John Havens	Integrated DNA Technologies	jhavens@idtdna.com		
35	Todd Smith	Geospiza, Inc	todd@geospiza.com		
36	Lori Peterson	Caldera Pharmaceuticals, Inc	court@cpsci.com		
37	Nicole Touchet	Caldera Pharmaceuticals, Inc	touchet@cpsci.com		
38 39	Jane Hutchinson David Graham	Roche Applied Science Oak Ridge National Laboratory	jane.hutchinson@roche.com grahamde@ornl.gov		
40	Dibyendu Kumar	University of Florida	dkumar@ufl.edu		
41	Cristina Vesbach	University of New Mexico	cvesbach@unm.edu, cvesbach@gmail.com		
42	Lydia Zeglin	University of New Mexico	Lydia.Zeglin@oregonstate.edu		
43	Dave Van Horn	University of New Mexico	vanhorn@unm.edu		
44	Jim Knight	Roche Diagnostics - 454	iames.knight@roche.com		
45	Paula Imbro	TAURI Group	paula.imbro@taurigroup.com		
46	John Schlager	AFRL	John.Schlager@WPAFB.AF.MIL		
47	Keven Stevens	Integrated DNA Technologies	kstevens@idtdna.com		
48	Folker Meyer	Argonne National Laboratory	folker@anl.gov		
49	Sepideh Rogers	Roche Applied Science	sepideh.rogers@roche.com		
50	Caleb Phillips	Texas Tech University	caleb.phillips@ttu.edu		
51	Darren Grafham	The Wellcome Trust Sanger Institute	dg1@sanger.ac.uk		
52	Kimberly Bishop-Lilly	Naval Medical Research Center	kim.bishop-lilly@med.navy.mil		
53	Mohit Patel	Naval Medical Research Center	Mohit.Patel@med.navy.mil		
54	Jim Bristow	Joint Genome Institute - LBL	JBristow@lbl.gov		
55	Roger Lasken	J Craig Venter Institute	rlasken@jcvi.org		
56	Dave Klaasse Mary Neu	DTRA - TMT	david.klaasse@DTRA.MIL		
57 58	Diane Dutt	Los Alamos National Laboratory DTRA-TMT	mneu@lanl.gov diane.dutt@dtra.mil		
59	John Oliver	NABsys Inc.	oliver@nabsys.com		
60	Faye Schilkey	National Center for Genome Resources (NCGR)	fds@ncgr.org		
61	Andrew Bradbury	Los Alamos National Laboratory	amb@lanl.gov		
62	David Bruce	Los Alamos National Laboratory	dbruce@lanl.gov		
63	Uffe Hellsten	Lawrence Berkeley National Laboratory	uhellsten@lbl.gov		
64	Cynthia Hendrickson	New England Biolabs, Inc.	hendrickson@neb.com		
65	Steve Turner	Pacific Biosciences	sturner@pacificbiosciences.com, trard@pacificbiosciences.com		
66	Gary Qiao	DTRA - CB	Guilin.Qiao@DTRA.MIL		
67	Sara D'angelo	Los Alamos National Laboratory	sdangelo@lanl.gov		
86	Greg Meyers	DTRA-TMT Tauri Group	gregory.meyers_anser@taurigroup.com		
69	Haley Fiske	Illumina, Inc.	hfiske@illumina.com		
70	Nileena Velappan	Los Alamos National Laboratory	nileena@lanl.gov		
71	Scott Remine	DTRA-TMT	scott.remine@dtra.mil		
72	Miranda Harmon-Smith	JGI - Lawrence Livermore National Laboratory	MLHarmon-Smith@lbl.gov		
73	Julia Scheerer	DTRA-TMT Tauri Group	Julia.Scheerer_CONTRACTOR@dtra.mil		
74	Chris Daum	JGI - Lawrence Livermore National Laboratory	daum1@llnl.gov		
75 76	Craig Blackhart	Los Alamos National Laboratory	blackhart@lanl.gov		
76 77	Riley Arnaudville Stacey Broomall	Los Alamos National Laboratory Edgewood Chemical Biological Center	rileya@lanl.gov stacey.broomall@us.army.mil		
78	Sofi Ibrahim	USAMRIID	stacey.broomaii@us.army.mii sofi.ibrahim@us.army.mil		
0					
79	Bharat Patel	Griffith University, Brisbane, Australia	b.patel@griffith.edu.au		

8-10 New York State Control of Co				
8.8 junes Debyeneds C. Lements Jedech Provid Laboratory (2009) Bit get (2009) Bit	81	Amy Powell	Sandia National Laboratory	ajpowel@sandia.gov
85 Beneau Designation 1 Charge Years of Lawrence Heroid National Justices on the Committee of C			·	
See Program Security Securi				
See Deseal Resource 1. Cristy Venter Institute (CVI) 2. Bourse Standard 3. Deseal Resource 3. See				
Beautiful Common Comm				
Section Comparison Compar				
Martine Stamiller Gil Lemmes Berkere Nistonal Laboratory				
Maile Young Sanda Marcon Laboratories mmy-rough anticle apreciation mmy-rough anticle	87	Nadia Fedorova		
92 Tit Response Control to Contro	88	Matthew Hamilton	JGI - Lawrence Berkeley National Laboratory	MGHamilton@lbl.gov
3 Gen Feetach Les American Laboratory Listen of Brid age	89	Malin Young	Sandia National Laboratories	mmyoung@sandia.gov
6 Gery Reside. 6 Gery Reside. 7 Gery Reside. 7 Gery Reside. 8 National Security of Comments of Comments of Comments. 8 National Management of Comments. 8 National Management. 9 National Manag	90	Jim Laugharn	Covaris, Inc.	JLaugharn@covarisinc.com
92 Gebrech May 93 Cheber May 94 Charles Among Annie Among Andrew Control Control 95 Cheber May 95 Charles Among Annie Among An				
Exception May Sends National Laboratories Sends National Lab				
9-1 (No. 1) Meng Sude DOS Distored Laboratories, Sirgapore 10, 2004, Lonc Bidson up a 10				
See Liver State DOD Prisonal Laboratories, Singapore Summary State on the private of the state of the				
South December Purdue University Internance Int				
Mail Lemon				
Mail Lemons Droof restatute Droof Restatut				
Symme Geoderin See Airmon National Laboratory JGI Symme See Airmon Symmetry State Louding Symmetry Sym				
Modern Verbrome Proc. Sanger Institute Modern Mod	98	Niall Lennon		nlennon@broadinstitute.org
September Westmington Liversity Strong Westmington Westmingt	99	Lynne Goodwin	Los Alamos National Laboratory - JGI	lynneg@lanl.gov
Vincent Magrint Visioning Durinorsity School magnetis wasten wasted du Majora Control Comprome Resources (NCCR) make floor or gray magnetis wasten wasted du National Control Comprome Resources (NCCR) make floor or gray mak	100	Andries Van Tonder	Wellcome Trust Sanger Institute	avt@sanger.ac.uk
Nyan Kim	101	Bob Fulton	Washington University in St. Louis	bfulton@watson.wustl.edu
Nyan Kim	102	Vincent Magrini	Washington University School	vmagrini@watson.wustl.edu
Sergey Koren J. Gray Verter Institute Sergey Koren D. Gray Verter Institute Sergey Koren D. Garrick Peters Serge Celling of Medicine Sergey Celli		-		· ·
Divisation Bushay Borytor College of Mendicine chushay@box.mic.com dipletes@Signorprent_com			,	
Common Degram Rocha Solytor College of Medicine stocked Both and				
1907 Shannon Dugan Rocha 1908 Yan Ding 1908 Yan Ding 1909 Libratine Sum 100 Christine Sum 100 Christin				
Teach Particles Sum U.C. Berkeiser Service College of Medicine Video States Sum U.C. Berkeiser Service States Sum Service States Sum Service States Sum Service States Stat				
Christine Sum University of Mison, Italy Street Mispone Christine Laun® ginnal com				
Christine Sum University of Mison, Italy Street Mispone Christine Laun® ginnal com				
Flavio Mignone University of Mian, Isay Sci Cancer Approx, Genome Sciences Centre navarann Biogoca	109	Christine Sun	UC Berkeley	
See Warren SC Cancer Agency, Genome Sciences Centre marten Bibogac, ca marten Bibogac				flavio.mignone@unimi.it
113 Tory Li Pitzer Tury				
1113 Tony LI Pitzer Pitzer ITL@wyeth.com Herby tetrauth Pitzer dieffroy dieffroy dieffroy tetrauth Pitzer dieffroy				
115 John Petrosino Baylor Collego of Medicine petrosino petrosino Baylor Collego of Medicine petrosino petrosi				
116 Avind Bharti The National Center for Genome Resources (NCGR) abb Bing org grid				
Tempor T				
117 Feng Chen Joint Genome Institute LBL Ficher@Bib.gov Individual Cartest for Genome Resources (NCGR) Individual Cartest for Genome Cartest for Cartest				J1
Tell Robin Kramer The National Center for Genome Resources (NCGR) mts@ncgr.org link@ncgr.org link@				
1190 Karen Davenport Los Alamos National Laboratory Kwdavenport® land gov				
Ruth Glidenthorp Wellcome Trust Sanger Institute mg 78 aanger a.u.k mg 78 aanger a.u.	118	Robin Kramer	The National Center for Genome Resources (NCGR)	rsk@ncgr.org
121 Lauren Robertson Welcome Trust Sanger Institute If 88 sanger ac.u.k	119	Karen Davenport	Los Alamos National Laboratory	kwdavenport@lanl.gov
123 Anne Babbage Vellcome Trust Sanger Institute abt @sanger acut k 213 Anne Babbage Vellcome Trust Sanger Institute abt @sanger acut k 214 Peter Vander Horn Life Technologies peter vander Horn @lifetech.com 125 Miles Fitzgerald Broad Institute of MIT drop with the peter vander from @lifetech.com 126 Chelsea Dumbar Broad Institute of MIT drop @broad mit.edu 127 Ann Montmayeur Broad Institute of MIT annamon@broad.mit.edu 128 Chris Miller U.C. Berkeley carniller@gmail.com 129 Donna Muzzry Baylor College of Medicine donname Born.tim.cedu 130 Michael Holder Baylor College of Medicine donname Born.tim.cedu 131 Rob Holt Brisis Columbia Cancer Agency Genome Sciences Centre miholder@bcm.tim.cedu 131 Rob Holt Brisis Columbia Cancer Agency Genome Sciences Centre miholder@bcm.tim.cedu 132 Jean Zhao LBNL/JG synhamolia Laboratory Lam@@lint.gov 133 Marisa Lam Lawrence Livemmor National Laboratory Lam@@lint.gov 134 Valerie Barbe Genoscope - CEA vipate@genoscope.cm.sfr 135 Eric Peletietr Genoscope - CEA vipate@genoscope.cm.sfr 136 Mary Campbell Los Alamos National Laboratory mcampbell@list.gov 137 Beverly Parson-Quintana Los Alamos National Laboratory mcampbell@list.gov 138 Kevin Henry Simon Fraser University kanney@stat.genoscope.cm.sfr 140 Teles Reserved ECEC cincle reserved @u.ca 141 Nacies Reserved ECEC cincle reserved @u.ca 142 Jil Bartfeld U.C Berkeley Derkeley Derkeley 143 Anamos Stational Laboratory bead @list.gov 144 Linda Melncke Los Alamos National Laboratory bead @list.gov 145 Academ Stational Laboratory bead @list.gov 146 Linda Melncke Los Alamos National Laboratory bead @list.gov 147 Linda Melncke Los Alamos National Laboratory bead @list.gov 148 Lamer Stational Laboratory bead @list.gov 149 Linda Melncke Los Alamos National Laboratory bead @list.go	120	Ruth Gilderthorp	Wellcome Trust Sanger Institute	rg7@sanger.ac.uk
123 Anne Babbage Vellcome Trust Sanger Institute abt @sanger acut k 213 Anne Babbage Vellcome Trust Sanger Institute abt @sanger acut k 214 Peter Vander Horn Life Technologies peter vander Horn @lifetech.com 125 Miles Fitzgerald Broad Institute of MIT drop with the peter vander from @lifetech.com 126 Chelsea Dumbar Broad Institute of MIT drop @broad mit.edu 127 Ann Montmayeur Broad Institute of MIT annamon@broad.mit.edu 128 Chris Miller U.C. Berkeley carniller@gmail.com 129 Donna Muzzry Baylor College of Medicine donname Born.tim.cedu 130 Michael Holder Baylor College of Medicine donname Born.tim.cedu 131 Rob Holt Brisis Columbia Cancer Agency Genome Sciences Centre miholder@bcm.tim.cedu 131 Rob Holt Brisis Columbia Cancer Agency Genome Sciences Centre miholder@bcm.tim.cedu 132 Jean Zhao LBNL/JG synhamolia Laboratory Lam@@lint.gov 133 Marisa Lam Lawrence Livemmor National Laboratory Lam@@lint.gov 134 Valerie Barbe Genoscope - CEA vipate@genoscope.cm.sfr 135 Eric Peletietr Genoscope - CEA vipate@genoscope.cm.sfr 136 Mary Campbell Los Alamos National Laboratory mcampbell@list.gov 137 Beverly Parson-Quintana Los Alamos National Laboratory mcampbell@list.gov 138 Kevin Henry Simon Fraser University kanney@stat.genoscope.cm.sfr 140 Teles Reserved ECEC cincle reserved @u.ca 141 Nacies Reserved ECEC cincle reserved @u.ca 142 Jil Bartfeld U.C Berkeley Derkeley Derkeley 143 Anamos Stational Laboratory bead @list.gov 144 Linda Melncke Los Alamos National Laboratory bead @list.gov 145 Academ Stational Laboratory bead @list.gov 146 Linda Melncke Los Alamos National Laboratory bead @list.gov 147 Linda Melncke Los Alamos National Laboratory bead @list.gov 148 Lamer Stational Laboratory bead @list.gov 149 Linda Melncke Los Alamos National Laboratory bead @list.go	121	Lauren Robertson	Wellcome Trust Sanger Institute	lr8@sanger.ac.uk
123 Anne Babbage Wellcome Trust Sanger Institute ab 1 @sanget acuk 124 Peter Vander Horn Life Technologies peter vanderhorn @lifetch com 125 Mike Fitzgerald Broad Institute of MIT disple Woodmitted or 126 Chiese Dunbar Broad Institute of MIT disple Woodmitted or 127 Anna Montralyeur Broad Institute of MIT annamon(@broad mit.edu 128 Christ Miller U.U Berkeley Sender Grand Institute of MIT annamon(@broad mit.edu 129 Donna Muzry Bayfor College of Medicine donnam@bcm.tmc.edu 130 Michael Holder Bayfor College of Medicine donnam@bcm.tmc.edu 131 Rob Holt British Columbia Cancer Agency Genome Sciences Centre rhotl@bogs.c.ca 132 Jean Zhao LBNU,Gi zyzhade/Bil.gov 133 Marisa Lam Lawrence Livermore National Laboratory Lam9@ind.gov 134 Valerie Barbe Genoscope - CEA sharbe@genoscope cns.fr 135 Eric Pelletier Genoscope - CEA dender @genoscope cns.fr 136 Mary Campbell Los Alamos National Laboratory bapq@lianl.gov 137 Beverly Parson-Quintana Los Alamos National Laboratory bapq@lianl.gov 138 Kein Henry Simon Fraser University bapq@lianl.gov 139 Jamie Scott Simon Fraser University kahenyi @fu.cs 140 Felix Breden Simon Fraser University kahenyi @fu.cs 141 Nicole Rosenzweig ECBC nicole Review simple Bil.gov 142 Jill Banfeld U.C Barkeley National Laboratory bapq@lianl.gov 143 Kanwar Singh JGI - Lawrence Berkeley National Laboratory singh@bil.gov 144 Lida Melincke Los Alamos National Laboratory singh@bil.gov 145 Joe Salvatore CLC Bio sharos National Laboratory singh@bil.gov 146 Right Melhole Los Alamos National Laboratory singh@bil.gov 147 Chris Munk Los Alamos National Laboratory mick@lianl.gov 148 Right Melhole Los Alamos National Laboratory mick@lianl.gov 149 Julia Banfeld U.C Barkeley Sinon Fraser University singh@bil.gov 149 Julia Banfeld Los Ala				
Peter Vander Horn Life Technologies peter vanderhorn @lifetech.com			·	
126 Mike Fitzgerald Broad Institute of MIT cloley@broadinistute org 127 Anna Montmayeur Broad Institute of MIT cloley@broadinistute org 128 Chris Miller UC Berkeley csmiller@gmail.com 129 Donna Muzny Bayfor College of Medicine donnam@broad.mit.edu 130 Michael Holder Bayfor College of Medicine donnam@broad.mit.edu 131 Rob Holt Bristh Columbia Cancer Agency Genome Sciences Centre rholf@bcgsc.ca 132 Jean Zhao LENLYJG zhyzho@ltls.gov 133 Marisa Lam Lawrence Livermore National Laboratory Lam@@linl.gov 134 Valerie Barbe Genoscope - CEA dharbe@genoscope.cns.fr 136 Eric Palletier Genoscope - CEA dharbe@genoscope.cns.fr 136 Eric Palletier Genoscope - CEA dharbe@genoscope.cns.fr 137 Beverly Parson-Quintana Los Alamos National Laboratory bapq@linl.gov 137 Beverly Parson-Quintana Los Alamos National Laboratory bapq@linl.gov 138 Kavin Henry Simon Fraser University shaheny@stu.ca 140 Felix Broden Simon Fraser University phedem@stu.ca 141 Nicole Rosenzweig ECBC nicolerosenzweig@us.amry.mil 142 Jill Banfield UC Berkeley Jill Jamine Jill Jamine Jill Jamine Jill Jamine 143 Kanwar Singh JG - Lawrence Berkeley National Laboratory singn@ibl.gov 144 Lild Meincke Los Alamos National Laboratory singn@ibl.gov 145 Joe Salvatore CLC Bio shahon Salmon Shational Laboratory singn@ibl.gov 146 Rod Meincke Los Alamos National Laboratory singn@ibl.gov 147 Chris Murik Los Alamos National Laboratory singn@ibl.gov 148 Atalwar Teshina Los Alamos National Laboratory singn@ibl.gov 149 Julia Meincke Los Alamos National Laboratory singn@ibl.gov 140 Felix Mirk Los Alamos National Laboratory clash@ianl.gov 141 Lidd Meincke Los Alamos National Laboratory singn@ibl.gov 142 Jill Banfield Los Alamos National Laboratory singn@ibl.gov clash@ianl.gov 145 Lidd Meincke Los Alamos				
126 Chelsea Dumbar Broad Institute of MIT International Colley@broadinstitute.org 172 Anna Montmayeur Broad Institute of MIT International Colley International College International Colle				
127 Anna Montmayeur Broad Institute of MIT Senkseley Camiller® gmail.com 128 Obna Muzry Baylor College of Medicine donnam@bcm.tmc.edu 139 Michael Holder Baylor College of Medicine donnam@bcm.tmc.edu 131 Rob Holt British Columbia Cancer Agency Genome Sciences Centre thot@bcgsc.ca 132 Jean Zhao LENLUJGI 133 Marisa Lam Lawrence Livermore National Laboratory Lamre@linl.gov 134 Valerie Barbe Genoscope − CEA Volarbe@genoscope.ors.fr 135 Eric Pelletier Genoscope − CEA Volarbe@genoscope.ors.fr 136 Eric Pelletier Genoscope − CEA delice genoscope.ors.fr 137 Bevery Person-Quintana Los Alamos National Laboratory bapq@lanl.gov 138 Kevin Henry Simon Fraser University kahenny@stu.ca 140 Felix Breden Simon Fraser University breden@stu.ca 141 Nicole Rosenzweig ECBC cincle Researce ECBC 142 Jill Banfield U.C Berkeley Boren College Bancer 143 Kawara Singh JG - Lawrence Berkeley National Laboratory bandled@berkeley.odu 144 Linda Meincke Los Alamos National Laboratory bandled@berkeley.odu 145 Los Salvatore C.C. Bio 146 Brittany Hold Los Alamos National Laboratory bandled@berkeley.odu 147 Chris Munk Los Alamos National Laboratory bandled@berkeley.odu 148 Razuwi Teshina Los Alamos National Laboratory bandled@berkeley.odu 149 Kyle Hubbard Los Alamos National Laboratory bandled@berkeley.odu 149 Kyle Hubbard Los Alamos National Laboratory bandled@berkeley.odu 149 Natural Salvatore C.C. Bio salvatore C.C. Bio 140 Salvatore C.C. Bio Salv				
128 Chris Miller U.C. Berkeley Camiller@gmail.com donnam@bom.tm.edu donnam.tm.edu donnam				
130 Michael Holder Baylor College of Medicine donnam@bom.tmc.edu donnam.college donnam.col				
130 Michael Holder Baylor College of Medicine mholder@bcm.tmc.edu 131 Rob Holt British Columbia Cancer Agency Genome Sciences Centre mhold@bcgc.ca 132 Jean Zhao LBNLJG zyzhao@lbl.gov 133 Marisa Lam Lawrence Livermore National Laboratory Lame@ilni.gov 134 Valerie Barbe Genoscope - CEA whatbe@genoscope.cns.fr 135 Eric Pelletier Genoscope - CEA encp@genoscope.cns.fr 136 Mary Campbell Los Alamos National Laboratory mcampbell@lanl.gov 137 Beverly Parson-Quintana Los Alamos National Laboratory bapg@lanl.gov 138 Koriw Henry Simon Fraser University kaheny@slu.ca 139 Jamie Scott Simon Fraser University kaheny@slu.ca 140 Felix Breden Simon Fraser University kaheny@slu.ca 141 Nicole Rosenzweig ECBC nicole-rosenzweig@us.amy.mil 142 Jill Banfeld UC Berkeley 143 Kanwar Singh J.Gl - Lawrence Berkeley National Laboratory sanytaron-Quintana 146 Brittary Held Los Alamos National Laboratory sanytaron-Quintana 147 Chris Munk Los Alamos National Laboratory sanytaron-Quintana 148 Hazuki Teshina Los Alamos National Laboratory sanytaron-Quintana 149 Kale Held Los Alamos National Laboratory sanytaron-Quintana 140 Lawrence Berkeley National Labor				
131 Rob Holt British Columbia Cancer Agency Genome Sciences Centre Inchil Blogsc.ca Inchil Blogsc.	129	Donna Muzny	Baylor College of Medicine	donnam@bcm.tmc.edu
132 Jean Zhao LBNLJ/G xyrhao@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Safet Sa	130	Michael Holder	Baylor College of Medicine	mholder@bcm.tmc.edu
132 Jean Zhao LBNLJ/G xyrhao@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Lm@@lbl.gov Safet Sa	131	Rob Holt	British Columbia Cancer Agency Genome Sciences Centre	rholt@bcgsc.ca
133	132			
134 Valerie Barbe Genoscope - CEA vbarbe@genoscope.cns.fr 135 Eric Pelletier Genoscope - CEA ericip@genoscope.cns.fr 136 Mary Campbell Los Alamos National Laboratory mcampbell@lanl.gov 137 Beverly Parson-Quintana Los Alamos National Laboratory bapq@lanl.gov 138 Kevin Henry Simon Fraser University kahenry@stlu.ca 139 Jamie Scott Simon Fraser University kahenry@stlu.ca 140 Felix Breden Simon Fraser University breden@stlu.ca 141 Nicole Rosenzweig ECBC nicole.rosenzweig@us.army.mil 142 Jill Banfield UC Berkeley banfield@berkeley.edu 143 Kanwar Singh J.Gl - Lawrence Berkeley National Laboratory ksingh@bl.gov 144 Linda Melincke Los Alamos National Laboratory meincke @lanl.gov 145 Joe Salvatore Cl.C Bio galvatore@clobio.com 146 Brittany Held Los Alamos National Laboratory bheld@lanl.gov 147 Chris Munik Los Alamos National Laboratory bheld@lanl.gov 148 Hazuki Teshima Los Alamos National Laboratory hazuki@lanl.gov 149 Kyle Hubbard US Army Edgewood Chemical Biological Center kyle.hubbard@us.army.mil 150 Janna Nichols Los Alamos National Laboratory mich@lanl.gov 151 Patti Wills Los Alamos National Laboratory mich@lanl.gov 152 Kim McMurry Los Alamos National Laboratory mich@lanl.gov 153 Leigh Cash Los Alamos National Laboratory mich@lanl.gov 154 Jannafier Harris Los Alamos National Laboratory mich@lanl.gov 155 Liz Saunders Los Alamos National Laboratory mich@lanl.gov 156 Lit Luo University of Texas Health Science Center at Houston Lituo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory dense@lanl.gov 158 Lituo University of Texas Health Science Center at Houston Lituo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory dense@lanl.gov 158 Robert Sinabaugh University of Texas Health Science Center at Houston Lituo@uth.tmc.edu 159 Graha				
135 Eric Pelletier Genoscope - CEA				
136 Mary Campbell Los Alamos National Laboratory bapq@lanl.gov				
137 Beverty Parson-Quintana Los Alamos National Laboratory bapq@lant.gov 138 Kevin Henry Simon Fraser University kahenry@stu.ca 139 Jamie Scott Simon Fraser University jkscott@stu.ca 140 Felix Breden Simon Fraser University breden@sfu.ca 141 Nicole Rosenzweig ECBC nicole.rosenzweig@us.army.mil 142 Jill Banfield UC Berkeley jbanfield@berkeley.edu 143 Kanwar Singh JGI - Lawrence Berkeley National Laboratory ksingh@bl.gov 144 Linda Meincke Los Alamos National Laboratory meincke@lant.gov 145 Joe Salvatore CLC Bio jsalvatore@clcbio.com 146 Brittany Held Los Alamos National Laboratory bheld@lant.gov 147 Chris Munk Los Alamos National Laboratory cmunk@lant.gov 148 Hazuki Teshima Los Alamos National Laboratory hazuki@lant.gov 149 Kyle Hubbard US Army Edgewood Chemical Biological Center kyle.hubbard@us.army.mil 150 Janna Nichols Los Alamos National Laboratory inich@lant.gov 151 Patti Wills Los Alamos National Laboratory wills@lant.gov 152 Kim McMurry Los Alamos National Laboratory inich@lant.gov 153 Leigh Cash Los Alamos National Laboratory kmcmurry@lant.gov 155 Liz Saunders Los Alamos National Laboratory kmcmurry@lant.gov 156 Lit Luo University of Texas Health Science Center at Houston LiLuo@uth.mc.edu 157 Olga Chertkov Los Alamos National Laboratory chefw@lant.gov 158 Graham Scott Life Technologies Graham.Scott@lifetech.com 159 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory Ukesros@open.com 161 Tanja Woyke JGI - Lawrence Berkeley National Laboratory Ukesros@open.com 162 Thomas Okerson OpGen twagner@open.com 163 Mike Fitzpatrick OpGen fitzpatrick@open.com 164 Trevor Wagner OpGen				
138 Kevin Henry Simon Fraser University Kahenry@stu.ca 139 Jamie Scott Simon Fraser University Recott@stu.ca 140 Felix Breden Simon Fraser University Breden@stu.ca 141 Nicole Rosenzweig ECBC Inicole.rosenzweig@us.army.mil 142 Jill Banfield UC Berkeley Berkeley edu 143 Kanwar Singh JGI - Lawrence Berkeley National Laboratory Isanifeld Berkeley.edu 144 Linda Meincke Los Alamos National Laboratory Resident 145 Joe Salvatore CLC Bio Salvatore CLC Bio 146 Brittany Held Los Alamos National Laboratory Berkeley Cichio.com 147 Chris Munk Los Alamos National Laboratory Berkeley Alamos National Laboratory 148 Hazuki Teshima Los Alamos National Laboratory Cmunk@lanl.gov 149 Kyle Hubbard US Army Edgewood Chemical Biological Center Ryle.hubbard@us.army.mil 150 Jana Nichols Los Alamos National Laboratory Inich@lanl.gov 151 Patti Wills Los Alamos National Laboratory Wills@lanl.gov 152 Kim McMurry Los Alamos National Laboratory Wills@lanl.gov 153 Leigh Cash Los Alamos National Laboratory Richmurry@lanl.gov 154 Jennifer Harris Los Alamos National Laboratory Richmurry@lanl.gov 155 Liz Saunders Los Alamos National Laboratory Richmurry@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li.Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory Graham Scott@lifetech.com 158 Graham Scott Life Technologies Graham Scott@lifetech.com 159 Graham Scott Life Technologies Graham Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory JLee2@lb.lgov 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lb.lgov 162 Thomas Okerson OpGen Independent Indepen				
139 Jamie Scott Simon Fraser University ijkscott@sfu.ca 140 Felix Breden Simon Fraser University breden@sfu.ca 141 Nicole Rosenzweig ECBC nicole.rosenzweig@us.army.mil 142 Jill Banfield U.C Berkeley dipanfield@berkeley.edu 143 Kanwar Singh JGI - Lawrence Berkeley National Laboratory ksingh@bl.gov 144 Linda Meincke Los Alamos National Laboratory meincke@lanl.gov 145 Joe Salvatore CLC Bio isalvatore@clcbio.com 146 Brittany Held Los Alamos National Laboratory bheld@lanl.gov 147 Chris Munk Los Alamos National Laboratory cmunk@lanl.gov 148 Hazuki Teshima Los Alamos National Laboratory hazuki@lanl.gov 149 Kyle Hubbard U.S Army Edgewood Chemical Biological Center kyle.hubbard@us.army.mil 150 Janna Nichols Los Alamos National Laboratory minc@lanl.gov 151 Patt Willis Los Alamos National Laboratory minc@lanl.gov 152 Kim McMurry Los Alamos National Laboratory wills@lanl.gov 153 Leigh Cash Los Alamos National Laboratory kmcmurry@lanl.gov 154 Jennifer Harris Los Alamos National Laboratory kmcmurry@lanl.gov 155 Liz Saunders Los Alamos National Laboratory fharris@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory drisinsab@unm.edu 158 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory Jue?@bl.gov 161 Tanja Woyke JGI - Lawrence Berkeley National Laboratory Jue?@bl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory Jue?@bl.gov 164 Trevor Wagner OpGen tokerson@opgen.com 165 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory Nickyrpides@bl.gov		•		
140 Felix Breden Simon Fraser University breden@sfu.ca 141 Nicole Rosenzweig ECBC nicole.rosenzweig@us.army.mil 142 Jill Banfield UC Berkeley Banfield@berkeley.edu 143 Kanwar Singh JGI - Lawrence Berkeley National Laboratory ksingh@lb.lg.ov 144 Linda Meincke Los Alamos National Laboratory meincke@lanl.gov 145 Joe Salvatore CLC Bio 146 Brittany Held Los Alamos National Laboratory bheld@lanl.gov 147 Chris Munk Los Alamos National Laboratory bheld@lanl.gov 148 Hazuki Teshima Los Alamos National Laboratory nazuki@lanl.gov 149 Kyle Hubbard US Army Edgewood Chemical Biological Center kyle.hubbard@us.army.mil 150 Janna Nichols Los Alamos National Laboratory mink@lanl.gov 151 Patti Wills Los Alamos National Laboratory wills@lanl.gov 152 Kim McMurry Los Alamos National Laboratory wills@lanl.gov 153 Leigh Cash Los Alamos National Laboratory loash@lanl.gov 154 Jennifer Harris Los Alamos National Laboratory loash@lanl.gov 155 Liz Saunders Los Alamos National Laboratory loash@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory ehs@lanl.gov 158 Robert Sinsabaugh University of New Mexico fishsab@unm.edu 159 Graham Scott Life Technologies Graham.Scott@ifetech.com 150 Tanja Woyke JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory Dobgrides@bl.gov 164 Trevor Wagner OpGen tokerson@opgen.com 165 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@bl.gov				
141 Nicole Rosenzweig ECBC				
142 Jill Banfield UC Berkeley				
143 Kanwar Singh JGI - Lawrence Berkeley National Laboratory ksingh@lbl.gov 144 Linda Meincke Los Alamos National Laboratory meincke@lanl.gov 145 Joe Salvatore CLC Bio jsalvatore@clcbio.com 146 Brittany Held Los Alamos National Laboratory bheld@lanl.gov 147 Chris Munk Los Alamos National Laboratory cmunk@lanl.gov 148 Hazukil Teshima Los Alamos National Laboratory hazuki@lanl.gov 149 Kyle Hubbard US Army Edgewood Chemical Biological Center kyle.hubbard@us.army.mil 150 Janna Nichols Los Alamos National Laboratory inich@lanl.gov 151 Patti Wills Los Alamos National Laboratory inich@lanl.gov 152 Kim McMurry Los Alamos National Laboratory wills@lanl.gov 153 Leigh Cash Los Alamos National Laboratory kmcmurry@lanl.gov 154 Jannifer Harris Los Alamos National Laboratory icash@lanl.gov 155 Liz Saunders Los Alamos National Laboratory fiharris@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li.Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory ochritx@lanl.gov 158 Robert Sinsabaugh University of New Mexico risinsab@unm.edu 159 Graham Scott Life Technologies Graham.Scott@lifetech.com 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Mike Fitzpatrick OpGen miftzpatrick@opgen.com 164 Trevor Wagner OpGen tokerson@opgen.com 165 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@bil.gov 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@bil.gov 167 NCKyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@bil.gov 168 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@bil.gov NCKyrpides@bil.gov NCKyrpides@bil.gov NCKyrpides@bil.gov NCKyrpides Laboratory NCKyrpides Laboratory NCKyrpides Laborator				
144 Linda Meincke Los Alamos National Laboratory meincke@lanl.gov 145 Joe Salvatore CLC Bio jsalvatore@clcbio.com 146 Brittany Held Los Alamos National Laboratory bheld@lanl.gov 147 Chris Munk Los Alamos National Laboratory cmunk@lanl.gov 148 Hazuki Teshima Los Alamos National Laboratory hazuki@lanl.gov 149 Kyle Hubbard US Army Edgewood Chemical Biological Center kyle.hubbard@us.army.mil 150 Janna Nichols Los Alamos National Laboratory minch@lanl.gov 151 Patti Wills Los Alamos National Laboratory mils@lanl.gov 152 Kim McMurry Los Alamos National Laboratory mils@lanl.gov 153 Leigh Cash Los Alamos National Laboratory kmcmurry@lanl.gov 154 Jennifer Harris Los Alamos National Laboratory leash@lanl.gov 155 Liz Saunders Los Alamos National Laboratory ifharris@lanl.gov 155 Liz Saunders Los Alamos National Laboratory ehs@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li.Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory ochtkv@lanl.gov 158 Robert Sinsabaugh University of New Mexico fisinsab@umm.edu 159 Graham Scott Life Technologies Graham Scott Life Technol				
145 Joe Salvatore CLC Bio ijsalvatore@clcbio.com ijsalvatore@clcbio.com 146 Brittany Held Los Alamos National Laboratory bheld@lanl.gov cmunk@lanl.gov munk@lanl.gov 147 Chris Munk Los Alamos National Laboratory cmunk@lanl.gov hazuki@lanl.gov hazuki@lanl.gov hazuki@lanl.gov hazuki@lanl.gov land laboratory hazuki@lanl.gov hazuki@lanl.gov land laboratory hazuki@lanl.gov hazuki@lanl.gov land laboratory linich@lanl.gov				
146 Brittany Held	144		Los Alamos National Laboratory	
146 Brittany Held	145	Joe Salvatore	CLC Bio	jsalvatore@clcbio.com
Chris Munk	146	Brittany Held	Los Alamos National Laboratory	bheld@lanl.gov
Hazuki Teshima				
149 Kyle Hubbard US Army Edgewood Chemical Biological Center kyle.hubbard@us.army.mil 150 Janna Nichols Los Alamos National Laboratory inich@lanl.gov inich@lanl.gov 151 Patti Wills Los Alamos National Laboratory wills@lanl.gov wills@lanl.gov Wills@lanl.gov kmcmurry@lanl.gov Los Alamos National Laboratory Ifharris@lanl.gov 154 Jennifer Harris Los Alamos National Laboratory ifharris@lanl.gov 155 Liz Saunders Los Alamos National Laboratory ehs@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li.Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory ochrtkv@lanl.gov 158 Robert Sinsabaugh University of New Mexico Ifsinsab@unm.edu 159 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen Itolamos Okerson OpGen				
150				
Total Wills				
152 Kim McMurry Los Alamos National Laboratory kmcmurry@lanl.gov 153 Leigh Cash Los Alamos National Laboratory lcash@lanl.gov 154 Jennifer Harris Los Alamos National Laboratory jfharris@lanl.gov 155 Liz Saunders Los Alamos National Laboratory ehs@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li.Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory ochrtkv@lanl.gov 158 Robert Sinsabaugh University of New Mexico risinsab@unm.edu 159 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Mike Fitzpatrick OpGen mfitzpatrick@opgen.com 164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov NCKyrpides@lbl.gov NCKyrpides@lbl.gov NCKyrpides@lbl.gov NCKyrpides@lbl.gov 165 Nicos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
153				
154				
155 Liz Saunders Los Alamos National Laboratory ehs@lanl.gov 156 Li Luo University of Texas Health Science Center at Houston Li.Luo@uth.tmc.edu 157 Olga Chertkov Los Alamos National Laboratory ochtkv@lanl.gov 158 Robert Sinsabaugh University of New Mexico rlsinsab@unm.edu 159 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory TWoyke@lbl.gov 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Mike Fitzpatrick OpGen mfitzpatrick@opgen.com 164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
156				
157 Olga Chertkov				
158 Robert Sinsabaugh University of New Mexico Irlsinsab@unm.edu 159 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory TWoyke@lbl.gov 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Mike Fitzpatrick OpGen mfitzpatrick@opgen.com 164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
159 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory TWoyke@lbl.gov 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Mike Fitzpatrick OpGen mfitzpatrick@opgen.com 164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov	157	Olga Chertkov		ochrtkv@lanl.gov
159 Graham Scott Life Technologies Graham.Scott@lifetech.com 160 Tanja Woyke JGI - Lawrence Berkeley National Laboratory TWoyke@lbl.gov 161 Janey Lee JGI - Lawrence Berkeley National Laboratory JLee2@lbl.gov 162 Thomas Okerson OpGen tokerson@opgen.com 163 Mike Fitzpatrick OpGen mfitzpatrick@opgen.com 164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
Tanja Woyke JGI - Lawrence Berkeley National Laboratory TWoyke@lbl.gov				
161				
162 Thomas Okerson OpGen tokerson@opgen.com 163 Mike Fitzpatrick OpGen mfitzpatrick@opgen.com 164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@ibl.gov				
163 Mike Fitzpatrick OpGen mfitzpatrick@opgen.com 164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
164 Trevor Wagner OpGen twagner@opgen.com 165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
165 Robert Edgar Independent bob@drive5.com 166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
166 Nikos Kyrpides JGI - Lawrence Berkeley National Laboratory NCKyrpides@lbl.gov				
		Nikos Kyrpides	JGI - Lawrence Berkeley National Laboratory	NCKyrpides@lbl.gov
167 Wes Warren The Genome Center/Washington University School of Medicine wwarren@watson.wustl.edu	166			

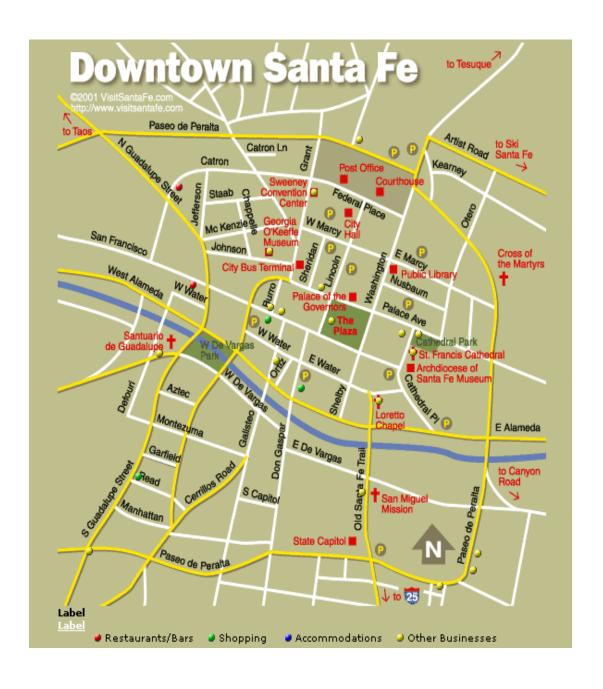
400	I=: 0	T-1 0 0 1 M 1: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
168	Tina Graves	The Genome Center/Washington University School of Medicine	tgraves@watson.wustl.edu
169	Patrick Minx	The Genome Center/Washington University School of Medicine	pminx@watson.wustl.edu
170	Catrina Fronick	The Genome Center/Washington University School of Medicine	cstrowma@watson.wustl.edu
171	Wendy Schackwitz	JGI - Lawrence Berkeley National Laboratory	WSSchackwitz@lbl.gov
172	Jeff Martin	JGI - Lawrence Berkeley National Laboratory	JAMartin@lbl.gov
173	Joel Martin	JGI - Lawrence Berkeley National Laboratory	J_Martin@lbl.gov
174	Zhong Wang	JGI - Lawrence Berkeley National Laboratory	ZhongWang@lbl.gov
175	Vito Mangiardi	JGI - Lawrence Berkeley National Laboratory	VJMangiardi@lbl.gov
176	Susannah Tringe	JGI - Lawrence Berkeley National Laboratory	GSTringe@lbl.gov
177	Massie Ballon	JGI - Lawrence Berkeley National Laboratory	MLBallon@lbl.gov
178	David Gilbert	JGI - Lawrence Berkeley National Laboratory	DEGilbert@lbl.gov
179	Tijana Glavina del Rio	JGI - Lawrence Berkeley National Laboratory	TGlavina@lbl.gov
180	Alla Lapidus	JGI - Lawrence Berkeley National Laboratory	ALapidus@lbl.gov
181	Ed Kirton	JGI - Lawrence Berkeley National Laboratory	ESKirton@lbl.gov
182	Stephan Trong	JGI - Lawrence Livermore National Laboratory	STrong@lbl.gov
183	James Han	JGI - Lawrence Berkeley National Laboratory	jkhan@lbl.gov
184	Zhili He	University of Oklahoma	zhili.he@ou.edu
185	Brian Foster	JGI - Lawrence Berkeley National Laboratory	Bfoster@lbl.gov
186	Steve Lowry	JGI - Lawrence Berkeley National Laboratory	SRLowry@lbl.gov
187	Alicia Clum	JGI - Lawrence Berkeley National Laboratory	Aclum@lbl.gov
188	Hui Sun	JGI - Lawrence Berkeley National Laboratory	Hsun@lbl.gov
189	Sarah Young	The Broad Institute	stowey@broadinstitute.org
190	David Heiman	The Broad Institute	dheiman@broadinstitute.org
191	Terry Shea	The Broad Institute	tshea@broadinstitute.org
191	Dan Conway	CLC Bio	dconway@clcbio.com
192		CLC Bio	
	Saul Kravitz	Los Alamos National Laboratory	skravitz@clcbio.com han cliff@lanl.gov
194	Cliff Han	·	
195	Quynh Doan	Life Technologies	Quynh.Doan@lifetech.com
196	Ben Allen	Los Alamos National Laboratory	bsa@lanl.gov
197	Ayush Ramana	University of Pittsburgh (UPMC)	ramana2@upmc.edu
198	Helen Cui	Los Alamos National Laboratory	hhcui@lanl.gov
199	Kurt LaButti	JGI - Lawrence Berkeley National Laboratory	klabutti@lbl.gov
200	Andy Seirp	Los Alamos National Laboratory	aseirp@lanl.gov
201	Dariusz Przybylski	Broad Institute of MIT	dariusz@broadinstitute.org
202	Ahmet Zeytun	Los Alamos National Laboratory	azeytun@lanl.gov
203	Alex Quezada	Los Alamos National Laboratory	alexq@lanl.gov
204	Lucy Zhang	Los Alamos National Laboratory	xlz@lanl.gov
205	Hajni Daligault	Los Alamos National Laboratory	hajkis@lanl.gov
206	David Sims	HudsonAlpha Institute for Biotechnology	dsims@hudsonalpha.org
207	David L Hirschberg	Columbia University	david.hirschberg@columbia.edu
208	Rex Malmstrom	Lawrence Berkeley National Laboratory	RRMalmstrom@lbl.gov
209	Mark Roskey	Caliper Life Sciences	mark.roskey@caliperls.com
210	Stephanie Malfatti	JGI - Lawrence Livermore National Laboratory	malfatti3@llnl.gov
211	Neal Shea	Illumina, Inc.	nshea@illumina.com
212	Jon Armstrong	CoFactor Genomics	jon_armstrong@cofactorgenomics.com
213	Krista Reitenga	Los Alamos National Laboratory	reitenga@lanl.gov
214	Wei Gu	Los Alamos National Laboratory	gu_wei01@hotmail.com
215	Adi Beth-din	Israel Institute for Biological Research (IIRB)	adib@iibr.gov.il
216	Ofir Israeli	Israel Institute for Biological Research (IIRB)	ofiri@iibr.gov.il
217	Roxanne Tapia	Los Alamos National Laboratory	rox@lanl.gov
218	Joel Berendzen	Los Alamos National Laboratory	joelb@lanl.gov
219	Jon Suzuki	USDA-ARS-PBARC	Jon.Suzuki@ARS.USDA.GOV
220	Loren Hauser Darrell Dinwiddie	Oakridge National Laboratory The National Center for Genome Resources (NCGR)	hauserlj@ornl.gov
221			dld@ncgr.org
222	Shannon Hateley	The National Center for Genome Resources (NCGR)	slh@ncgr.org
223	Jennifer van Velkinburgh	The National Center for Genome Resources (NCGR)	jcv@ncgr.org
224	Take Ogawa	RainDance	OGAWAT@raindancetech.com
225	Keith Brown	RainDance	BROWNK@raindancetech.com
226	Lin Pham	RainDance	PHAML@raindancetech.com
227	Johar Ali	Ontario Institute for Cancer Research (OICR)	Johar.Ali@oicr.on.ca
228	Emily Zentz	OpGen	ezentz@opgen.com
229	Daniel Clements	HudsonAlpha Genome Sequencing Center	dclements@hudsonalpha.org
230	Nicole Schoenknecht	HudsonAlpha Genome Sequencing Center	nschoenknecht@hudsonalpha.org
231	Bette Korber	Los Alamos National Laboratory	btk@lanl.gov
232	Patrick Chain	Los Alamos National Laboratory	pchain@lanl.gov
233	Scott Geib	USDA-Agricultural Research Services	Scott.Geib@ARS.USDA.GOV
234	Saied Jaradat	Princess Haya Biotechnology Center, Jordan University of Science and	sjaradat@just.edu.jo
		Technology	
235	Manfred Lee	Life Technologies	Manfred.Lee@lifetech.com
236	Lynette Kizer	CME GROUP INC.	lkizer@cmegroup-inc.com
237	Tim Minogue	USAMRIID	timothy.minogue@us.army.mil
238	Steve Lott	Pacific Biosciences	slott@pacificbiosciences.com
239	Jason Roos	DTRA	jason.roos@us.army.mil
240	Maria Gomez	HudsonAlpha Genome Sequencing Center	mgomez@hudsonalpha.org; csmith@hudsonalpha.org
241	Joanna McGill	HudsonAlpha Genome Sequencing Center	jmcgill@hudsonalpha.org
242	Shijie Yao	CME GROUP INC.	syao@cmegroup-inc.com
243	Martin Hansen	University of Copenhagen	mail@maasha.dk
244	Peter Hraber	Los Alamos National Laboratory	phraber@lanl.gov
245	Bruce Jamieson	MICROSONIC SYSTEMS	bruce.jamieson@microsonics.com
240	Di ace varinesori	INIOTOGGINO OTOTENIO	Dragonjamicoonemicroscinics.com

		2	010 SFAF Attendee List - sorted	by NAME
Name (Sorte	d By Name)	FF#	Affiliation	email
Johar	Ali	227	Ontario Institute for Cancer Research (OICR	Johar.Ali@oicr.on.ca
Ben	Allen	196	Los Alamos National Laboratory	bsa@lanl.gov
Jon	Armstrong	212	CoFactor Genomics	jon_armstrong@cofactorgenomics.com
Riley	Arnaudville	76	Los Alamos National Laboratory	rileya@lanl.gov
Anne	Babbage	123		ab1@sanger.ac.uk
			Wellcome Trust Sanger Institute	
Massie	Ballon	177	JGI - Lawrence Berkeley National Laborator	MLBallon@lbl.gov
Jill	Banfield	142	UC Berkeley	jbanfield@berkeley.edu
Valerie	Barbe	134	Genoscope – CEA	vbarbe@genoscope.cns.fr
Tara	Bennink	13	Edge BioSystems	TBennink@edgebio.com
Joel	Berendzen	218	Los Alamos National Laboratory	joelb@lanl.gov
Adi	Beth-din	215	Israel Institute for Biological Research (IIRB	adib@iibr.gov.il
Arvind	Bharti	116	The National Center for Genome Resources (NCGR)	akb@ncgr.org
Alistar	Bishop	29	Chemical and Biological Detection Group, Defence Science and Technology Laboratory (DSTL) Porton Dowr	DWCLEARY@mail.dstl.gov.uk
Kimberly	Bishop-Lilly	52	Naval Medical Research Center	kim.bishop-lilly@med.navy.mil
Craig	Blackhart	75	Los Alamos National Laboratory	blackhart@lanl.gov
Robert	Blakesley	15	NIH Intramural Sequencing Center (NISC)	rblakesl@nhgri.nih.gov
Andrew	Bradbury	61	Los Alamos National Laboratory	amb@lanl.gov
Felix	Breden	140	Simon Fraser University	breden@sfu.ca
Jim	Bristow	54		JBristow@lbl.gov
			Joint Genome Institute - LBI	
Shelise	Brooks	16	NIH-NHGRI-NISC	sbrooks@mail.nih.gov
Stacey	Broomall	77	Edgewood Chemical Biological Center	stacey.broomall@us.army.mil
Keith	Brown	225	RainDance	BROWNK@raindancetech.com
David	Bruce	62	Los Alamos National Laboratory	dbruce@lanl.gov
Christian	Buhay	105	Baylor College of Medicine	cbuhay@bcm.tmc.edu
Mary	Campbell	136	Los Alamos National Laboratory	mcampbell@lanl.gov
Leigh	Cash	153	Los Alamos National Laboratory	Icash@lanl.gov
Patrick	Chain	232	Los Alamos National Laboratory	pchain@lanl.gov
Jean	Challacombe	80	Los Alamos National Laboratory	jchalla@lanl.gov
			`	
Feng	Chen	117	Joint Genome Institute - LBL	FChen@lbl.gov
Olga	Chertkov	157	Los Alamos National Laboratory	ochrtkv@lanl.gov
Daniel	Clements	229	HudsonAlpha Genome Sequencing Center	dclements@hudsonalpha.org
Alicia	Clum	187	JGI - Lawrence Berkeley National Laborator	Aclum@lbl.gov
			· · · · · · · · · · · · · · · · · · ·	-
Holly	Coleman	14	The National Institutes of Health Intramural Sequencing Center (NISC)	hcoleman@mail.nih.gov
Dan	Conway	192	CLC Bio	dconway@clcbio.com
Helen	Cui	198	Los Alamos National Laboratory	hhcui@lanl.gov
Hajni	Daligault	205	Los Alamos National Laboratory	hajkis@lanl.gov
Sara	D'angelo	67	Los Alamos National Laboratory	sdangelo@lanl.gov
Chris	Daum	74	JGI - Lawrence Livermore National Laboratory	daum1@llnl.gov
Karen	Davenport	119	Los Alamos National Laboratory	kwdavenport@lanl.gov
Shweta	Deshpande	83	JGI - Lawrence Berkeley National Laborator	SDeshpande@lbl.gov
Chris	Detter	1	Los Alamos National Laboratory - JG	cdetter@lanl.gov
Janine	Detter	11	Los Alamos National Laboratory	Janined@lanl.gov
			,	
Armand	Dichosa	21	Los Alamos National Laboratory	armand@lanl.gov
Yan	Ding	108	Baylor College of Medicine	yding@bcm.tmc.edu
Darrell	Dinwiddie	221	The National Center for Genome Resources (NCGR)	dld@ncgr.org
Quynh	Doan	195	Life Technologies	Quynh.Doan@lifetech.com
Shannon	Dugan-Rocha	107	Baylor College of Medicine	sdugan@bcm.edu
Chelsea	Dunbar	126	Broad Institute of MIT	cfoley@broadinstitute.org
Diane	Dutt	58	DTRA-TMT	diane.dutt@dtra.mil
Robert	Edgar	165	Independent	bob@drive5.com
Nadia	Fedorova	87		NFedorova2@jcvi.org
Haley	Fiske	69	Illumina, Inc.	hfiske@illumina.com
Mike	Fitzgerald	125	Broad Institute of MIT	fitz@broad.mit.edu
Mike	Fitzpatrick	163	OpGen	mfitzpatrick@opgen.com
Michael	Fitzsimons	31	Los Alamos National Laboratory	msfitz@lanl.gov
Brian	Foster	185	JGI - Lawrence Berkelev National Laborator	Bfoster@lbl.gov
Claire	Fraser-Liggett	30	Institute for Genome Sciences, University of Maryland School of Medicine	
Kenneth	Frey	17	Airforce Research Laboratory	Kenneth.Frey@WPAFB.AF.MIL
Catrina	Fronick	170	The Genome Center/Washington University School of Medicine	cstrowma@watson.wustl.edu
Bob	Fulton	101	Washington University in St. Louis	bfulton@watson.wustl.edu
Tiziano	Gaiotto	91	Los Alamos National Laboratory	tiziano@lanl.gov
Scott	Geib	233	USDA-Agricultural Research Services	Scott.Geib@ARS.USDA.GOV
David	Gilbert	178	JGI - Lawrence Berkeley National Laborator	DEGilbert@lbl.gov
Ruth	Gilderthorp	120	Wellcome Trust Sanger Institute	rg7@sanger.ac.uk
	Glavina del Rio	179	JGI - Lawrence Berkeley National Laboratory	TGlavina@lbl.gov
Tijana	Gomez	240	HudsonAlpha Genome Sequencing Center	mgomez@hudsonalpha.org; csmith@hudsonalpha.org
Tijana Maria	GUITIEZ		Los Alamos National Laboratory - JG	lynneg@lanl.gov
Maria	Goodwin	99		
Maria Lynne	Goodwin		The Wellcome Trust Sanger Institute	Ida1@sanger.ac.uk
Maria Lynne Darren	Goodwin Grafham	51	The Wellcome Trust Sanger Institute Oak Ridge National Laboratory	dg1@sanger.ac.uk
Maria Lynne Darren David	Goodwin Grafham Graham	51 39	Oak Ridge National Laboratory	grahamde@ornl.gov
Maria Lynne Darren David Tina	Goodwin Grafham Graham Graves	51 39 168	Oak Ridge National Laboratory The Genome Center/Washington University School of Medicine	grahamde@ornl.gov tgraves@watson.wustl.edu
Maria Lynne Darren David Tina Wei	Goodwin Grafham Graham Graves Gu	51 39 168 214	Oak Ridge National Laboratory The Genome Center/Washington University School of Medicine Los Alamos National Laboratory	grahamde@ornl.gov tgraves@watson.wustl.edu gu_wei01@hotmail.com
Maria Lynne Darren David Tina Wei Matthew	Goodwin Grafham Graham Graves Gu Hamilton	51 39 168 214 88	Oak Ridge National Laboratory The Genome Center/Washington University School of Medicine Los Alamos National Laboratory JGI - Lawrence Berkeley National Laborator	grahamde@ornl.gov tgraves@watson.wustl.edu gu_wei01@hotmail.com MGHamilton@lbl.gov
Maria Lynne Darren David Tina Wei Matthew James	Goodwin Grafham Graham Graves Gu Hamilton Han	51 39 168 214 88 183	Oak Ridge National Laboratory The Genome Center/Washington University School of Medicine Los Alamos National Laboratory JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Berkeley National Laboratory	grahamde@ornl.gov tgraves@watson.wustl.edu gu_wei01@hotmail.com MGHamilton@lbl.gov jkhan@lbl.gov
Maria Lynne Darren David Tina Wei Matthew James Cliff	Goodwin Grafham Graham Graves Gu Hamilton Han	51 39 168 214 88 183 194	Oak Ridge National Laboratory The Genome Center/Washington University School of Medicine Los Alamos National Laboratory JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Berkeley National Laboratory Los Alamos National Laboratory	grahamde@ornl.gov tgraves@watson.wustl.edu gu_wei01@hotmail.com MGHamilton@lbl.gov jkhan@lbl.gov han_cliff@lanl.gov
Maria Lynne Darren David Tina Wei Matthew James Cliff Martin	Goodwin Grafham Graham Graves Gu Hamilton Han Han	51 39 168 214 88 183 194 243	Oak Ridge National Laboratory The Genome Center/Washington University School of Medicine Los Alamos National Laboratory JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Berkeley National Laboratory Los Alamos National Laboratory University of Copenhagen	grahamde@ornl.gov tgraves@watson.wustl.edu gu_wei01@hotmail.com MGHamilton@lbl.gov jkhan@lbl.gov han_cliff@lanl.gov mail@maasha.dk
Maria Lynne Darren David Tina Wei Matthew James Cliff	Goodwin Grafham Graham Graves Gu Hamilton Han	51 39 168 214 88 183 194	Oak Ridge National Laboratory The Genome Center/Washington University School of Medicine Los Alamos National Laboratory JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Berkeley National Laboratory Los Alamos National Laboratory University of Copenhagen	grahamde@ornl.gov tgraves@watson.wustl.edu gu_wei01@hotmail.com MGHamilton@lbl.gov jkhan@lbl.gov han_cliff@lanl.gov

Shannon Hateley Loren Hauser John Havens Zhili He David Heiman	220	The National Center for Genome Resources (NCGR) Oakridge National Laboratory Integrated DNA Technologies	slh@ncgr.org hauserlj@ornl.gov jhavens@idtdna.com
John Havens Zhili He David Heiman	34	Integrated DNA Technologies	
Zhili He David Heiman			jhavens@idtdna.com
David Heiman	184		
	107	University of Oklahoma	zhili.he@ou.edu
	190	The Broad Institute	dheiman@broadinstitute.org
Brittany Held	146	Los Alamos National Laboratory	bheld@lanl.gov
Uffe Hellsten	n 63	Lawrence Berkeley National Laborator	uhellsten@lbl.gov
Cynthia Hendric	ckson 64	New England Biolabs, Inc.	hendrickson@neb.com
Kevin Henry	138	Simon Fraser University	kahenry@sfu.ca
David Hirschb		Columbia University	david.hirschberg@columbia.edu
Michael Holder	130	Baylor College of Medicine	mholder@bcm.tmc.edu
Rob Holt	131	British Columbia Cancer Agency Genome Sciences Centre	rholt@bcgsc.ca
Jessica Hostetle		J Craig Venter Institute (JCVI)	Jessicah@jcvi.org
Peter Hraber	244	Los Alamos National Laboratory	phraber@lanl.gov
Kyle Hubbard		US Army Edgewood Chemical Biological Center	kyle.hubbard@us.army.mil
Tim Hunkap		Discovery Bio	tim@discoverybio.com
Jane Hutchin		Roche Applied Science	jane.hutchinson@roche.com
Sofi Ibrahim		USAMRIID	sofi.ibrahim@us.army.mil
Paula Imbro	45	TAURI Group	paula.imbro@taurigroup.com
Ofir Israeli	216	Israel Institute for Biological Research (IIRB	ofiri@iibr.gov.il
Bruce Jamieso	OII 245	MICROSONIC SYSTEMS	bruce.jamieson@microsonics.com
Saied Jaradat	234	Princess Haya Biotechnology Center, Jordan University of Science and	sjaradat@just.edu.jo
Oh ann an Ialana		Technology	
Shannon Johnson		Los Alamos National Laboratory	shannonj@lanl.gov
Ryan Kim	103	National Center for Genome Resources (NCGR)	rwk@ncgr.org
Stephen Kingsm		National Center for Genome Resources (NCGR)	sfk@ncgr.org
Ed Kirton	181	JGI - Lawrence Berkeley National Laboraton	ESKirton@lbl.gov
Lynette Kizer	236	CME GROUP INC.	lkizer@cmegroup-inc.com
Dave Klaasse		DTRA - TMT	david.klaasse@DTRA.MIL
Jim Knight	44	Roche Diagnostics - 454	james.knight@roche.com
Bette Korber	231	Los Alamos National Laboratory	btk@lanl.gov
Sergey Koren	104	J. Craig Venter Institute	skoren@jcvi.org
Robin Kramer		The National Center for Genome Resources (NCGR)	rsk@ncgr.org
Saul Kravitz	193	CLC Bio	skravitz@clcbio.com
Dibyendu Kumar	40	University of Florida	dkumar@ufl.edu
Nikos Kyrpide	es 166	JGI - Lawrence Berkeley National Laboraton	NCKyrpides@lbl.gov
Kurt LaButti	199	JGI - Lawrence Berkeley National Laborator	klabutti@lbl.gov
Marisa Lam	133	Lawrence Livermore National Laboratory	Lam9@linl.gov
Alla Lapidus	s 180	JGI - Lawrence Berkeley National Laborator	ALapidus@lbl.gov
Roger Lasken	55	J Craig Venter Institute	rlasken@jcvi.org
Jim Laughar	rn 90	Covaris, Inc.	JLaugharn@covarisinc.com
Janey Lee	161	JGI - Lawrence Berkeley National Laborator	JLee2@lbl.gov
Manfred Lee	235	Life Technologies	Manfred.Lee@lifetech.com
Niall Lennon		Broad Institute	nlennon@broadinstitute.org
Stan Letovsk		Independent	letovsky@bu.edu
Tony Li	113	Pfizer	TLi@wyeth.com
Koh Li Meng		DSO National Laboratories, Singapore	klimeng@dso.org.sg
Alfredo Lopez D		Novozymes, Inc.	ALLO@novozymes.com
Steve Lott	238	Pacific Biosciences	slott@pacificbiosciences.com
Steve Loury	186	JGI - Lawrence Berkeley National Laborator	SRLowry@lbl.gov
Li Luo	156	University of Texas Health Science Center at Houston	Li.Luo@uth.tmc.edu
Vincent Magrini		Washington University School	vmagrini@watson.wustl.edu
Stephanie Malfatti		JGI - Lawrence Livermore National Laboratory	malfatti3@llnl.gov
Rex Malmstr		Lawrence Berkeley National Laboratory	RRMalmstrom@lbl.gov
Vito Mangiar		JGI - Lawrence Berkeley National Laborator	VJMangiardi@lbl.gov
		IOL I B I I N C II I I	
	172 173	JGI - Lawrence Berkeley National Laborator	JAMartin@lbl.gov
Joel Martin Eric Mathur		JGI - Lawrence Berkeley National Laborator	J_Martin@lbl.gov EMathur@SyntheticGenomics.com
		Synthetic Genomics, Inc National Center for Genome Resources (NCGR)	
Greg May	5		gdm@ncgr.org
Elebeoba May Joanna McGill	93 241	Sandia National Laboratories HudsonAlpha Genome Sequencing Center	eemay@sandia.gov jmcgill@hudsonalpha.org
			mcgiii@nudsonaipna.org kmcmurry@lanl.gov
		Los Alamos National Laboratory	Isaac.Meek@caliperIs.com
Isaac Meek Linda Meincke	e 2	Caliper Life Sciences	meincke@lanl.gov
	48	Los Alamos National Laboratory	folker@anl.gov
		Argonne National Laboratory	
Greg Meyers		DTRA-TMT Tauri Group	gregory.meyers_anser@taurigroup.com
Flavio Mignone		University of Milan, Italy	flavio.mignone@unimi.it
Chris Miller	128	UC Berkeley	csmiller@gmail.com
Tim Minogue		USAMRIID	timothy.minogue@us.army.mil
Patrick Minx	169	The Genome Center/Washington University School of Medicine	pminx@watson.wustl.edu
Anna Montma		Broad Institute of MIT	annamont@broad.mit.edu
Joann Mudge	6	National Center for Genome Resources (NCGR)	jm@ncgr.org
Teri Mueller		Roche Diagnostics	teri.mueller@roche.com
Chris Munk	147	Los Alamos National Laboratory	cmunk@lanl.gov
Donna Muzny	129	Baylor College of Medicine	donnam@bcm.tmc.edu
Don Natvig	82	University of New Mexico	dnatvig@gmail.com
Mary Neu	57	Los Alamos National Laboratory	mneu@lanl.gov
Janna Nichols		Los Alamos National Laboratory	jnich@lanl.gov
Chad Nusbau		Broad Institute/MIT	chad@broadinstitute.org
Take Ogawa	224	RainDance	OGAWAT@raindancetech.com
Thomas Okersor		OpGen	tokerson@opgen.com
John Oliver	59	NABsys Inc.	oliver@nabsys.com
	-Quintana 137	Los Alamos National Laboratory	bapq@lanl.gov
Beverly Parson-			
	53	Naval Medical Research Center	Mohit.Patel@med.navy.mil
Beverly Parson-		Naval Medical Research Centel Griffith University, Brisbane, Australia	Mohit.Patel@med.navy.mil b.patel@griffith.edu.au
Beverly Parson- Mohit Patel	53 79		

			T	
Garrick	Peters	106	Ion Torrent	gpeters@iontorrent.com
Lori	Peterson	36	Caldera Pharmaceuticals, Inc	court@cpsci.com
Joe	Petrosino	115	Baylor College of Medicine	jpetrosi@bcm.tmc.edu
Lin	Pham	226	RainDance	PHAML@raindancetech.com
Caleb	Phillips	50	Texas Tech University	caleb.phillips@ttu.edu
Lee	Poeppelman	20	Airforce Research Laboratory	Lee.Poeppelman@WPAFB.AF.MIL
Amy	Powell	81	Sandia National Laboratory	ajpowel@sandia.gov
Dariusz	Przybylski	201	Broad Institute of MIT	dariusz@broadinstitute.org
Marcella	Putman	26	Life Technologies	Marcella.Putman@lifetech.com
Gary	Qiao	66	DTRA - CB	Guilin.Qiao@DTRA.MIL
Alex	Quezada	203	Los Alamos National Laboratory	alexq@lanl.gov
Diana	Radune	86	J Craig Venter Institute	DRadune@jcvi.org
Ayush	Ramana	197	University of Pittsburgh (UPMC)	ramana2@upmc.edu
Krista	Reitenga	213	Los Alamos National Laboratory	reitenga@lanl.gov
Scott	Remine	71	DTRA-TMT	scott.remine@dtra.mil
Gary	Resnick	92	Los Alamos National Laboratory	resnick@lanl.gov
Adam	Resnick	97	J Craig Venter Institute	aresnick@jcvi.org
Ernie	Retzel	9	National Center for Genome Resources (NCGR)	efr@ncgr.org
Michael	Rhodes	25	Life Technologies	Michael.Rhodes@lifetech.com
Lauren	Robertson	121	Wellcome Trust Sanger Institute	lr8@sanger.ac.uk
Sepideh	Rogers	49	Roche Applied Science	sepideh.rogers@roche.com
Jason	Roos	239	DTRA	jason.roos@us.army.mil
Nicole	Rosenzweig	141	ECBC	nicole.rosenzweig@us.army.mil
Mark	Roskey	209	Caliper Life Sciences	mark.roskey@caliperls.com
Joe	Salvatore	145	CLC Bio	jsalvatore@clcbio.com
Liz	Saunders	155	Los Alamos National Laboratory	ehs@lanl.gov
Wendy	Schackwitz	171	JGI - Lawrence Berkeley National Laborator	WSSchackwitz@lbl.gov
Julia	Scheerer	73	DTRA-TMT Tauri Group	Julia.Scheerer_CONTRACTOR@dtra.mil
Faye	Schilkey	60	National Center for Genome Resources (NCGR)	fds@ncgr.org
John	Schlager	46	AFRL	John.Schlager@WPAFB.AF.MIL
Nicole	Schoenknecht	230	HudsonAlpha Genome Sequencing Center	nschoenknecht@hudsonalpha.org
Jamie	Scott	139	Simon Fraser University	jkscott@sfu.ca
Graham	Scott	159	Life Technologies	Graham.Scott@lifetech.com
Andy	Seirp	200	Los Alamos National Laboratory	aseirp@lanl.gov
Terry	Shea	191	The Broad Institute	tshea@broadinstitute.org
Neal	Shea	211	Illumina, Inc.	nshea@illumina.com
Lou	Sherman	96	Purdue University	Isherman@purdue.edu
David	Sims	206	HudsonAlpha Institute for Biotechnology	dsims@hudsonalpha.org
Kanwar	Singh	143	JGI - Lawrence Berkeley National Laborator	ksingh@lbl.gov
Robert	Sinsabaugh	158	University of New Mexico	rlsinsab@unm.edu
Todd	Smith	35	Geospiza, Inc	todd@geospiza.com
Ng	Sock Hoon	94	DSO National Laboratories, Singapore	ng_sock_hoon@dso.org.sg
Keven	Stevens	47	Integrated DNA Technologies	kstevens@idtdna.com
Fiona	Stewart	18	New England Biolabs	stewart@neb.com
Christine	Sun	109	UC Berkeley	christine.l.sun@gmail.com
Hui	Sun	188	JGI - Lawrence Berkeley National Laborator	Hsun@lbl.gov
Jon	Suzuki	219	USDA-ARS-PBARC	Jon.Suzuki@ARS.USDA.GOV
Frank	Tansley	10	Life Technologies	Frank.Tansley@lifetech.com
Roxanne	Tapia	217	Los Alamos National Laboratory	rox@lanl.gov
Ken	Taylor	32	Integrated DNA Technologies	ktaylor@idtdna.com
Clotilde	Teiling			
Hazuki		28	Roche Applied Science	clotilde.teiling@roche.com
	Teshima	28 148	Los Alamos National Laboratory	clotilde.teiling@roche.com hazuki@lanl.gov
Jeff	Tetrault	28 148 114	Los Alamos National Laborator, Pfizer	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com
Jeff Brian	Tetrault Thomas	28 148 114 122	Los Alamos National Laborator, Pfizer UC Berkeley	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu
Jeff Brian Nicole	Tetrault Thomas Touchet	28 148 114 122 37	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com
Jeff Brian Nicole Susannah	Tetrault Thomas Touchet Tringe	28 148 114 122 37 176	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov
Jeff Brian Nicole Susannah Stephan	Tetrault Thomas Touchet Tringe Trong	28 148 114 122 37 176 182	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator, JGI - Lawrence Livermore National Laborator,	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov
Jeff Brian Nicole Susannah Stephan Steve	Tetrault Thomas Touchet Tringe Trong Turner	28 148 114 122 37 176 182 65	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator JGI - Lawrence Livermore National Laborator, Pacific Biosciences	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com
Jeff Brian Nicole Susannah Stephan Steve Peter	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn	28 148 114 122 37 176 182 65	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator, JGI - Lawrence Livermore National Laborator, Pacific Biosciences Life Technologies	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com
Jeff Brian Nicole Susannah Stephan Steve Peter Dave	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn	28 148 114 122 37 176 182 65 124	Los Alamos National Laboraton Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboraton JGI - Lawrence Livermore National Laboraton Pacific Biosciences Life Technologies University of New Mexicc	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Tonder	28 148 114 122 37 176 182 65 124 43	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator, JGI - Lawrence Livermore National Laborator, Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhom@lifetech.com vanhom@unm.edu avt@sanger.ac.uk
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Tonder van Velkinburgh	28 148 114 122 37 176 182 65 124 43 100 223	Los Alamos National Laboraton, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboraton JGI - Lawrence Livermore National Laboraton Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR)	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Tonder van Velkinburgh Velappan	28 148 114 122 37 176 182 65 124 43 100 223 70	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator JGI - Lawrence Livermore National Laborator Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR Los Alamos National Laborator)	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTring@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhom@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Tonder van Velkinburgh Velappan Vesbach	28 148 114 122 37 176 182 65 124 43 100 223 70 41	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator JGI - Lawrence Livermore National Laborator Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laborator, University of New Mexicc	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Tonder van Velkinburgh Velappan Vesbach Wagner	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator, JGI - Lawrence Livermore National Laborator, Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laborator, University of New Mexicc OpGen	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bothomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Velkinburgh Velappan Vesbach Wagner Walters	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator, JGI - Lawrence Livermore National Laborator, Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR Los Alamos National Laborator, University of New Mexicc OpGen Pacific Northwest National Laboratory	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhom@lifetech.com vanhom@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Yelkinburgh Velappan Vesbach Wagner Walters Wang	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174	Los Alamos National Laboraton Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboraton JGI - Lawrence Livermore National Laboraton Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laborator, University of New Mexicc University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboraton	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTring@lbl.gov Strong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@umm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Tonder van Velkinburgh Velappan Vesbach Wagner Walters Wang Warren	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111	Los Alamos National Laborator, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laborator JGI - Lawrence Livermore National Laborator Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR Los Alamos National Laborator, University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laborator BC Cancer Agency, Genome Sciences Centre	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Yelkinburgh Velappan Vesbach Wagner Walters Wang Warren	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@nogr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@bcgsc.ca
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Velkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhom@lifetech.com vanhom@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Rafal	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Tonder van Velkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills Woycicki	28 148 114 1122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTring@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ta.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov rafalwoycicki@gmail.com
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Rafal Tanja	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Yelkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills Woycicki Woyke	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24 160	Los Alamos National Laboraton, Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboraton JGI - Lawrence Livermore National Laboraton Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laborator, University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboraton BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laborator, Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laboraton	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov rafalwoycicki@gmail.com TWoyke@lbl.gov
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Rafal Tanja Shijie	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Yelkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills Woycicki Woyke Yao	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24 160 242	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laborator	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov rafalwoycicki@gmail.com TVoyke@lbl.gov syao@cmegroup-inc.com
Jeff Brian Nicole Susannah Stephan Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Rafal Tanja Shijie Malin	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Tonder van Velkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills Woycicki Woyke Yao Young	28 148 114 1122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24 160 242 89	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laborator CME GROUP INC. Sandia National Laboratories	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov Strong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov warren@bcgsc.ca wwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov rafalwoycicki@gmail.com TWoyke@lbl.gov syao@cmegroup-inc.com mmyoung@sandia.gov
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Rafal Tanja Shijie Malin Sarah	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Yonder Van Velkinburgh Velappan Vesbach Wagner Walters Wang Warren Wills Woycicki Woyke Yao Young	28 148 114 112 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24 160 242 89 189	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laborator CME GROUP INC. Sandia National Laboratories The Broad Institute	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov vwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov rafalwoycicki@gmail.com TWoyke@lbl.gov syao@cmegroup-inc.com mmyoung@sandia.gov stowey@broadinstitute.org
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Rafal Tanja Shijie Malin Sarah Lydia	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Tonder van Velkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills Woycicki Woyke Yao Young Zeglin	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 112 174 111 167 151 24 160 242 89 189 42	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laborator CME GROUP INC. Sandia National Laboratories The Broad Institute University of New Mexicc	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov rafalwoycicki@gmail.com TWoyke@lbl.gov syao@cmegroup-inc.com mmyoung@sandia.gov stowey@broadinstitute.org Lydia.Zeglin@oregonstate.edu
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Tanja Shijie Malin Sarah Lydia Emily	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Yelkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills Woycicki Woyke Yao Young Zeglin Zentz	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24 160 242 89 189 42 228	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laborator CME GROUP INC. Sandia National Laboratories The Broad Institute University of New Mexicc OpGen	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov stafalwoycicki@gmail.com TVoyke@lbl.gov syao@cmegroup-inc.com mmyoung@sandia.gov stowey@broadinstitute.org Lydia.Zeglin@oregonstate.edu ezentz@opgen.com
Jeff Brian Nicole Susannah Stephan Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Rafal Tranja Shijje Malin Sarah Lydia Emily Ahmet	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Tonder van Velkinburgh Velappan Vesbach Wagner Walters Wang Warren Wills Woycicki Woyke Yao Young Zeglin Zentz Zeytun	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24 160 242 89 189 42 228 202	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR) Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laborator CME GROUP INC. Sandia National Laboratories The Broad Institute University of New Mexicc OpGen Los Alamos National Laboratory	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov Strong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhom@lifetech.com vanhom@unm.edu avt@sanger.ac.uk jcv@nogr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov rafalwoycicki@gmail.com TWoyke@lbl.gov syao@cmegroup-inc.com mmyoung@sandia.gov stowey@broadinstitute.org Lydia.Zeglin@oregonstate.edu ezentz@opgen.com azeytun@lanl.gov
Jeff Brian Nicole Susannah Stephan Steve Peter Dave Andries Jennifer Nileena Cristina Trevor Ron Zhong Rene Wes Patti Tanja Shijie Malin Sarah Lydia Emily	Tetrault Thomas Touchet Tringe Trong Turner Vander Horn Van Horn Van Yelkinburgh Velappan Vesbach Wagner Walters Wang Warren Warren Wills Woycicki Woyke Yao Young Zeglin Zentz	28 148 114 122 37 176 182 65 124 43 100 223 70 41 164 12 174 111 167 151 24 160 242 89 189 42 228	Los Alamos National Laboratory Pfizer UC Berkeley Caldera Pharmaceuticals, Inc JGI - Lawrence Berkeley National Laboratory JGI - Lawrence Livermore National Laboratory Pacific Biosciences Life Technologies University of New Mexicc Wellcome Trust Sanger Institute The National Center for Genome Resources (NCGR Los Alamos National Laboratory University of New Mexicc OpGen Pacific Northwest National Laboratory JGI - Lawrence Berkeley National Laboratory BC Cancer Agency, Genome Sciences Centre The Genome Center/Washington University School of Medicine Los Alamos National Laboratory Warsaw University of Life Sciences - SGGW JGI - Lawrence Berkeley National Laborator CME GROUP INC. Sandia National Laboratories The Broad Institute University of New Mexicc OpGen	clotilde.teiling@roche.com hazuki@lanl.gov jeffrey.tetrault@pfizer.com bcthomas@berkeley.edu touchet@cpsci.com GSTringe@lbl.gov STrong@lbl.gov sturner@pacificbiosciences.com, trard@pacificbiosciences.com peter.vanderhorn@lifetech.com vanhorn@unm.edu avt@sanger.ac.uk jcv@ncgr.org nileena@lanl.gov cvesbach@unm.edu, cvesbach@gmail.com twagner@opgen.com ra.walters@pnl.gov ZhongWang@lbl.gov rwarren@bcgsc.ca wwarren@watson.wustl.edu wills@lanl.gov stafalwoycicki@gmail.com TVoyke@lbl.gov syao@cmegroup-inc.com mmyoung@sandia.gov stowey@broadinstitute.org Lydia.Zeglin@oregonstate.edu ezentz@opgen.com

Map of Santa Fe, NM



History of Santa Fe, NM

Thirteen years before Plymouth Colony was settled by the Mayflower Pilgrims, Santa Fe, New Mexico, was established with a small cluster of European type dwellings. It would soon become the seat of power for the Spanish Empire north of the Rio Grande. Santa Fe is the oldest capital city in North America and the oldest European community west of the Mississippi.

While Santa Fe was inhabited on a very small scale in 1607, it was truly settled by the conquistador Don Pedro de Peralta in 1609-1610. Santa Fe is the site of both the oldest public building in America, the Palace of the Governors and the nation's oldest community celebration, the Santa Fe Fiesta, established in 1712 to commemorate the Spanish reconquest of New Mexico in the summer of 1692. Peralta and his men laid out the plan for Santa Fe at the base of the Sangre de Cristo Mountains on the site of the ancient Pueblo Indian ruin of Kaupoge, or "place of shell beads near the water."

The city has been the capital for the Spanish "Kingdom of New Mexico," the Mexican province of Nuevo Mejico, the American territory of New Mexico (which contained what is today Arizona and New Mexico) and since 1912 the state of New Mexico. Santa Fe, in fact, was the first foreign capital over taken by the United States, when in 1846 General Stephen Watts Kearny captured it during the Mexican-American War.

Santa Fe's history may be divided into six periods:

Preconquest and Founding (circa 1050 to 1607)

Santa Fe's site was originally occupied by a number of Pueblo Indian villages with founding dates from between 1050 to 1150. Most archaeologists agree that these sites were abandoned 200 years before the Spanish arrived. There is little evidence of their remains in Santa Fe today.

The "Kingdom of New Mexico" was first claimed for the Spanish Crown by the conquistador Don Francisco Vasques de Coronado in 1540, 67 years before the founding of Santa Fe. Coronado and his men also discovered the Grand Canyon and the Great Plains on their New Mexico expedition.

Don Juan de Onate became the first Governor-General of New Mexico and established his capital in 1598 at San Juan Pueblo, 25 miles north of Santa Fe. When Onate retired, Don Pedro de Peralta was appointed Governor-General in 1609. One year later, he had moved the capital to present day Santa Fe.

Settlement Revolt & Reconquest (1607 to 1692)

For a period of 70 years beginning the early 17th century, Spanish soldiers and officials, as well as Franciscan missionaries, sought to subjugate and convert the Pueblo Indians of the region. The indigenous population at the time was close to 100,000 people, who spoke nine basic languages and lived in an estimated 70 multi-storied adobe towns (pueblos), many of which exist today. In 1680, Pueblo Indians revolted against the estimated 2,500 Spanish colonists in New Mexico, killing 400 of them and driving the rest back into Mexico. The conquering Pueblos sacked Santa Fe and burned most of the buildings, except the Palace of the Governors. Pueblo Indians occupied Santa Fe until 1692, when Don Diego de Vargas reconquered the region and entered the capital city after a bloodless siege.

Established Spanish Empire (1692 to 1821)

Santa Fe grew and prospered as a city. Spanish authorities and missionaries - under pressure from constant raids by nomadic Indians and often bloody wars with the Comanches, Apaches and Navajos-formed an alliance with Pueblo Indians and maintained a successful religious and civil policy of peaceful coexistence. The Spanish policy of closed empire also heavily influenced the lives of most Santa Feans during these years as trade was restricted to Americans, British and French.

The Mexican Period (1821 to 1846)

When Mexico gained its independence from Spain, Santa Fe became the capital of the province of New Mexico. The Spanish policy of closed empire ended, and American trappers and traders moved into the region. William Becknell opened the 1,000-mile-long Santa Fe Trail, leaving from Arrow Rock, Missouri, with 21 men and a pack train of goods. In those days, aggressive Yankeetraders used Santa Fe's Plaza as a stock corral. Americans found Santa Fe and New Mexico not as exotic as they'd thought. One traveler called the region the "Siberia of the Mexican Republic."

For a brief period in 1837, northern New Mexico farmers rebelled against Mexican rule, killed the provincial governor in what has been called the Chimayó Rebellion (named after a village north of Santa Fe) and occupied the capital. The insurrectionists were soon defeated, however, and three years later, Santa Fe was peaceful enough to see the first planting of cottonwood trees around the Plaza.

Territorial Period (1846 to 1912)

On August 18, 1846, in the early period of the Mexican American War, an American army general, Stephen Watts Kearny, took Santa Fe and raised the American flag over the Plaza. Two years later, Mexico signed the Treaty of Guadalupe Hidalgo, ceding New Mexico and California to the United States.

In 1851, Jean B. Lamy, arrived in Santa Fe. Eighteen years later, he began construction of the

Saint Francis Cathedral. Archbishop Lamy is the model for the leading character in Willa Cather's book, "Death Comes for the Archbishop."

For a few days in March 1863, the Confederate flag of General Henry Sibley flew over Santa Fe, until he was defeated by Union troops. With the arrival of the telegraph in 1868 and the coming of the Atchison, Topeka and the Santa Fe Railroad in 1880, Santa Fe and New Mexico underwent an economic revolution. Corruption in government, however, accompanied the growth, and President Rutherford B. Hayes appointed Lew Wallace as a territorial governor to "clean up New Mexico." Wallace did such a good job that Billy the Kid threatened to come up to Santa Fe and kill him. Thankfully, Billy failed and Wallace went on to finish his novel, "Ben Hur," while territorial Governor.

Statehood (1912 to present)

When New Mexico gained statehood in 1912, many people were drawn to Santa Fe's dry climate as a cure for tuberculosis. The Museum of New Mexico had opened in 1909, and by 1917, its Museum of Fine Arts was built. The state museum's emphasis on local history and native culture did much to reinforce Santa Fe's image as an "exotic" city.

Throughout Santa Fe's long and varied history of conquest and frontier violence, the town has also been the region's seat of culture and civilization. Inhabitants have left a legacy of architecture and city planning that today makes Santa Fe the most significant historic city in the American West.

In 1926, the Old Santa Fe Association was established, in the words of its bylaws, "to preserve and maintain the ancient landmarks, historical structures and traditions of Old Santa Fe, to guide its growth and development in such a way as to sacrifice as little as possible of that unique charm born of age, tradition and environment, which are the priceless assets and heritage of Old Santa Fe."

Today, Santa Fe is recognized as one of the most intriguing urban environments in the nation, due largely to the city's preservation of historic buildings and a modern zoning code, passed in 1958, that mandates the city's distinctive Spanish-Pueblo style of architecture, based on the adobe (mud and straw) and wood construction of the past. Also preserved are the traditions of the city's rich cultural heritage which helps make Santa Fe one of the country's most diverse and fascinating places to visit.

Go the Distance





Ultra high-fidelity synthesis of oligos up to 200 bases with mass spec QC!

IDT continues to push the boundaries of DNA synthesis quality and performance. By leveraging our industry-leading synthesis platform and chemistry, we now offer the custom synthesis of UltramersTM, 60-200 base DNA oligos suitable for demanding applications such as cloning and gene construction. Save time and trouble with direct synthesis of the entire target fragment!

To maintain our commitment to high quality, a proprietary LC-MS method has also been developed to provide accurate mass assessment for Ultramers TM . As always, this service is offered free of charge for each oligo with all data accessible online.

Ultramers[™] **specifications**:

- 60 to 200 DNA bases
- Delivered normalized and lyophilized in tubes
- Ideal for gene construction, cloning and ddRNAi
- PAGE purification and plate synthesis options also available



INNOVATION AND PRECISION IN NUCLEIC ACID SYNTHESIS





Order online by 2:00 P.M. ET and receive your SameDay® Oligos the next business morning via priority shipping.

- 2-OD guarantee (sufficient for > 250 PCR reactions)
- Deprotected & desalted
- Lyophilized
- Available within the U.S. and Canada
- 15-45 bases

Order today at www.idtdna.com



INNOVATION AND PRECISION IN NUCLEIC ACID SYNTHESIS





http://www.roche-diagnostics.us/

Meet and Greet Party

http://www.lifetech nologies.com Happy Hour x2 technologies™



http://www.caliperls.com/ Cooler Bags

> http://www.raindance technologies.com Lunch





http://www.illumina.com/

http://www.idtdna.com/ Home/Home.aspx Meeting Guides

